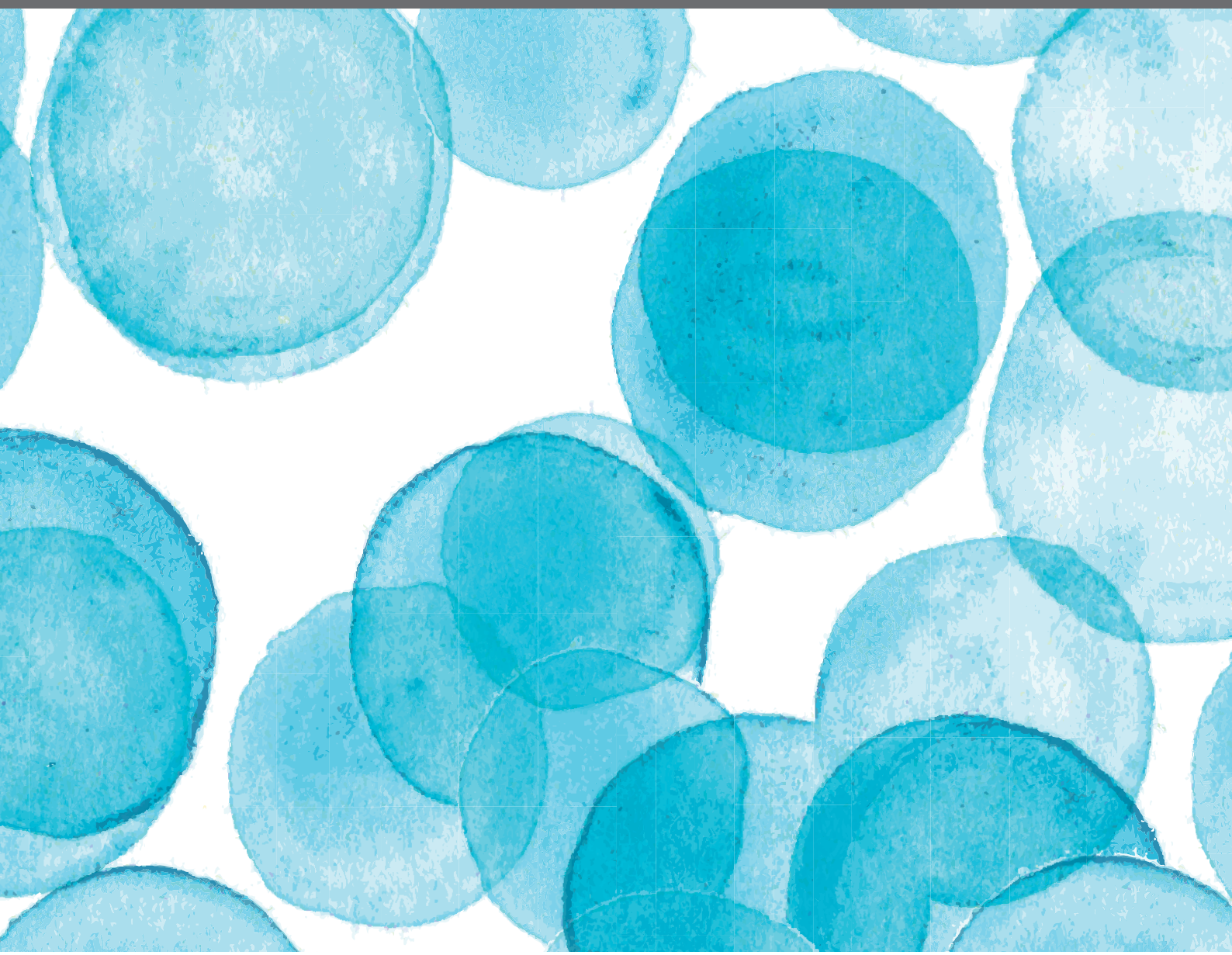
LARGE DNA VIRUSES – RESEARCH IN THIS OPENING DECADE

EDITED BY: Julien Andreani, Jonatas Abrahao and Masaharu Takemura
PUBLISHED IN: Frontiers in Microbiology



frontiers Research Topics

ISSN 1664-8714
ISBN 978-2-88976-752-6
DOI 10.3389/978-2-88976-752-6

About Frontiers

Frontiers is more than just an open-access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers Journal Series

The Frontiers Journal Series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the Frontiers Journal Series operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to Quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews.
Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the Frontiers Journals Series: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area! Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers Editorial Office: frontiersin.org/about/contact

LARGE DNA VIRUSES – RESEARCH IN THIS OPENING DECADE

Topic Editors:
Julien Andreani, Centre Hospitalier Universitaire de Grenoble, France
Jonatas Abrahao, Federal University of Minas Gerais, Brazil
Masaharu Takemura, Tokyo University of Scienceo, Japan

Citation: Andreani, J., Abrahao, J., Takemura, M., eds. (2022). Large DNA Viruses – Research in This Opening Decade. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-88976-752-6

Table of Contents

frontiers in Microbiology

ORIGINAL RESEARCH
published: 15 March 2021
doi: 10.3389/fmicb.2021.657471

Comparative Genomics and Environmental Distribution of Large dsDNA Viruses in the Family *Asfarviridae*

*Sangita Karki, Mohammad Moniruzzaman and Frank O. Aylward**

Department of Biological Sciences, Virginia Tech, Blacksburg, VA, United States

The family *Asfarviridae* is a group of nucleo-cytoplasmic large DNA viruses (NCLDVs) of which African swine fever virus (ASFV) is well-characterized. Recently the discovery of several *Asfarviridae* members other than ASFV has suggested that this family represents a diverse and cosmopolitan group of viruses, but the genomics and distribution of this family have not been studied in detail. To this end we analyzed five complete genomes and 35 metagenome-assembled genomes (MAGs) of viruses from this family to shed light on their evolutionary relationships and environmental distribution. The Asfarvirus MAGs derive from diverse marine, freshwater, and terrestrial habitats, underscoring the broad environmental distribution of this family. We present phylogenetic analyses using conserved marker genes and whole-genome comparison of pairwise average amino acid identity (AAI) values, revealing a high level of genomic divergence across disparate Asfarviruses. Further, we found that *Asfarviridae* genomes encode genes with diverse predicted metabolic roles and detectable sequence homology to proteins in bacteria, archaea, and eukaryotes, highlighting the genomic chimerism that is a salient feature of NCLDV. Our read mapping from Tara oceans metagenomic data also revealed that three *Asfarviridae* MAGs were present in multiple marine samples, indicating that they are widespread in the ocean. In one of these MAGs we identified four marker genes with > 95% AAI to genes sequenced from a virus that infects the dinoflagellate *Heterocapsa circularisquama* (HcDNAV). This suggests a potential host for this MAG, which would thereby represent a reference genome of a dinoflagellate-infecting giant virus. Together, these results show that *Asfarviridae* are ubiquitous, comprise similar sequence divergence as other NCLDV families, and include several members that are widespread in the ocean and potentially infect ecologically important protists.

Keywords: *Asfarviridae*, NCLDV, Megavirales, eukaryotic viruses, *Nucleocytoviricota*

OPEN ACCESS

Edited by:
Masaharu Takemura,
Tokyo University of Science, Japan

Reviewed by:
Thomas Klose,
Purdue University, United States
Keizo Nagasaki,
Usa Marine Biological Institute, Japan

****Correspondence:***
Frank O. Aylward
faylward@vt.edu

Specialty section:
This article was submitted to Virology, a section of the journal Frontiers in Microbiology

Received: *23 January 2021*
Accepted: *22 February 2021*
Published: *15 March 2021*

Citation:
Karki S, Moniruzzaman M and Aylward FO (2021) Comparative Genomics and Environmental Distribution of Large dsDNA Viruses in the Family Asfarviridae. Front. Microbiol. 12:657471. doi: 10.3389/fmicb.2021.657471

INTRODUCTION

The nucleo-cytoplasmic large DNA viruses (NCLDVs), also called *Nucleocytoviricota*, comprise a phylum of dsDNA viruses that infect diverse eukaryotes (Van Etten et al., 2010b; Koonin et al., 2020). NCLDVs include the largest viruses known, both in terms of virion size and genome length, and genomes within this group often contain genes involved in metabolic pathways

that are otherwise present only in cellular lineages (Fischer et al., 2010; Van Etten et al., 2010b; Schvarcz and Steward, 2018; Moniruzzaman et al., 2020a). Some families of NCLDV such as the *Poxviridae*, *Asfarviridae*, *Iridoviridae*, and *Phycodnaviridae* have been studied for decades, while others, such as the *Pandoraviridae*, *Mimiviridae*, and *Marseilleviridae*, have been discovered relatively recently (Raoult et al., 2004; Boyer et al., 2009; Philippe et al., 2013; Abergel et al., 2015). Although amoebae have been used as an effective system to cultivate many NCLDV, recent cultivation-independent studies have discovered a wide range of these viruses in diverse environments, suggesting that uncultivated members of this viral phylum are ubiquitous in the biosphere and infect diverse hosts (Monier et al., 2008; Hingamp et al., 2013; Bäckström et al., 2019; Endo et al., 2020; Moniruzzaman et al., 2020a; Schulz et al., 2020). Given the notable complexity of NCLDVs and their cosmopolitan distribution, there is a need to better understand their genomic diversity and biogeography.

The *Asfarviridae* is a family of NCLDVs for which the most well-studied member is the African swine fever virus (ASFV), an emerging pathogen that was first discovered in 1921 (Montgomery and Eustace Montgomery, 1921). Although ASFV has been extensively studied due to its high mortality rate and subsequent economic toll on livestock production, other viruses within the same family have remained relatively underexplored, and until recently ASFV was the only known member of the *Asfarviridae* family. In 2009, a virus infecting the marine dinoflagellate *Heterocapsa circularisquama* (HcDNAV) was cultivated, and partial sequencing of the DNA polymerase type B and MutS genes revealed that the virus likely belonged to the *Asfarviridae* (Ogata et al., 2009). Furthermore, a new amoeba virus, Faustovirus and other isolates of amoeba-infecting Asfarviruses that clustered with the *Asfarviridae* have also been reported (Reteno et al., 2015; Benamar et al., 2016). Using amoeba as the host, two other *Asfarviridae*, Kaumoebavirus and Pacmanvirus, were isolated (Bajrai et al., 2016; Andreani et al., 2017). Lastly, a culture independent study in early 2020 reported Asfar-like virus (AbalV) causing mass mortality in abalone (Matsuyama et al., 2020). Together, these studies have begun to show that the *Asfarviridae* are likely a diverse family of NCLDV that are globally distributed and infect both protist and metazoan hosts.

Recently, two studies (Moniruzzaman et al., 2020a; Schulz et al., 2020) reported numerous new metagenome-assembled genomes (MAGs) of NCLDV, some of which have phylogenetic affinity with the *Asfarviridae* family. However, the genomic characteristics of these MAGs have not been studied in detail. In this study, we leveraged five previously available Asfarvirus genomes and 35 new Asfarvirus MAGs to perform comparative genomic and biogeographic analysis of the *Asfarviridae* family and provide an assessment of the scale of Asfarvirus diversity in the environment. We assess the phylogenetic relationship of these new MAGs and previously discovered Asfarviruses to explore their evolutionary relationships, and we identify the potential evolutionary origins of the *Asfarviridae* genomic repertoires. We also report numerous genes encoding for different functions including central amino acid metabolism, nutrient homeostasis, and host infection. Moreover, we assess the distribution of marine Asfarvirus genomes in the ocean, and we identified high sequence similarity between one marine Asfarvirus MAG to marker gene sequences available from a virus known to infect the dinoflagellate *Heterocapsa circularisquama*, suggesting a potential host for this MAG. Our findings reveal that the *Asfarviridae* members are widespread in the ocean and potentially have roles in biogeochemical cycling through infection of ecologically-important protist lineages.

MATERIALS AND METHODS

Comparative Analysis and Protein Annotation

For this study, we analyzed 35 Asfarvirus MAGs generated in two previous studies (Moniruzzaman et al., 2020a; Schulz et al., 2020) and complete genomes of five Asfarviruses (Reteno et al., 2015; Silva et al., 2015; Bajrai et al., 2016; Andreani et al., 2017; Matsuyama et al., 2020). MAGs were quality-checked using ViralRecall v. 2.0 (default parameters), with results manually inspected to ensure that no large non-NCLDV contigs were present (Aylward and Moniruzzaman, 2021). We used Seqkit v0.12.0 (Shen et al., 2016) for FASTA/Q file manipulation to generate the statistics of the genomes and proteins. To predict protein and search for tRNA genes, we used Prodigal V2.6.3 (Hyatt et al., 2010) and ARAGORN v1.2.38 (Laslett, 2004), respectively, with default parameters. For the sequence similarity search, we used BLASTp against the NCBI reference sequence (RefSeq) database, version 92 (O'Leary et al., 2016). An E-value threshold of 1e-3 was used, and maximum target sequence was set to 1. Functional annotation of predicted proteins was done using hmmsearch (parameter -E 1e-5) in HMMER v3.3 (Eddy, 2011) against the EggNOG v.5 database (Huerta-Cepas et al., 2016) to assess the potential function of MAG-encoded proteins, and the best hits for each protein were recorded.

We calculated protein-level orthologous groups (OGs) shared between all genomes analyzed in this study using the Proteinortho tool version 6.0.14 (Lechner et al., 2011) with default parameters. The resulting matrix for the orthologous genes was used for the bipartite network analysis. A bipartite network for the 35 MAGs along with their reference genomes were constructed using igraph (Csardi and Nepusz, 2006), and selected members of *Poxviridae* were used as an outgroup. The network consisted of two node types, one for genomes and one for OGs. OGs that were present in at least one genome were analyzed. A Fruchterman-Reingold layout with 10,000 iteration was used for visualization purposes.

To assess the genomic diversity between Asfarviruses, we calculated amino acid identity (AAI) using the python script available at https://github.com/faylward/lastp_aai. This script uses LAST to detect bi-directional best hits to find the pairwise identity of orthologous proteins (Kiełbasa et al., 2011). The results were visualized using the gplots package (Warnes et al., 2020) in the R environment.

In order to assess the sequence similarity, the raw metagenomic reads from TARA ocean samples described

previously (Sunagawa et al., 2015) were downloaded from the NCBI SRA database, and forward Illumina reads were mapped against the selected genomes using LAST (Kiełbasa et al., 2011) with default parameters. The results were visualized with fragment recruitment plots using the ggplot2 package (Wickham, 2009) in the R environment.

Phylogenetic Reconstruction

To generate the phylogenetic tree, we analyzed 35 MAGs and five reference genomes along with selected members of the *Poxviridae* as an outgroup. We used five marker genes: major capsid protein (MCP), superfamily II helicase (SFII), virus-like transcription factor (VLTF3), DNA Polymerase B (PolB), and packaging ATPase (A32), that are previously shown to be useful and used for phylogenetic analysis of NCLDV MAGs (Yutin et al., 2009; Moniruzzaman et al., 2020a). We used a python script to identify the marker genes using hmmsearch (available at github.com/faylward/ncldv_markersearch), also previously described (Moniruzzaman et al., 2020a). We used Clustal Omega v1.2.4 (Sievers et al., 2011) for alignment, and trimAl v1.4.rev15 (Capella-Gutierrez et al., 2009) for alignment trimming (parameter -gt 0.1). We used IQ-TREE v. 1.6.12 (Minh et al., 2020) with the "-m TEST" model finder option (Kalyaanamoorthy et al., 2017) that identified VT+F+I+G4 as the best-fit model and 1,000 ultrafast bootstrap (Hoang et al., 2018) to reconstruct a maximum likelihood phylogenetic tree. Finally, we visualized the resulting phylogenetic tree using Interactive Tree of Life (iTOL) (Letunic and Bork, 2019).

Another phylogenetic tree was built using only PolB as a marker gene with the methods described previously. We did this because we observed that one NCLDV MAG (ERX552270.16) contained a PolB sequence with > 98% AAI to the PolB sequenced from the *Heterocapsa circularisquama* virus HcDNAV (Ogata et al., 2009) (as ascertained using BLASTP), and we wanted to confirm that these sequences clustered together. The complete genome of HcDNAV is not available, and so inclusion of this virus in the multi-locus tree was therefore not possible.

RESULTS AND DISCUSSION

Asfarvirus Genome Statistics

The Asfarvirus MAG assembly sizes ranged from 120 kbp (SRX802982.1) to 580.8 kbp (GVMAG-S-3300009702-144). Among the 35 MAGs, 17 had all five core genes used for phylogenetic analysis (A32, PolB, MCP, SFII, and VLTF3) while the rest of the genomes were missing only one core gene, including three MAGs in which the highly conserved PolB marker was not identified. This suggests that the MAGs are generally high quality, although the absence of some marker genes suggests that some are only nearly complete and that MAG assembly sizes are underestimates of the complete genome sizes. The % G+C content for the new MAGs ranged from 17 to 60%, while those of reference viruses ranged from 31 to 45%. The ARAGORN software predicted three tRNA genes (Leu, Ile, and Asn) for ERX552270.16, one Ile-tRNA gene for GVMAG-M-3300013133-40, GVMAG-M-3300023174-161, GVMAG-M-3300027793-10, GVMAG-S-3300005056-23, and GVMAG-S-3300010160-169, and one Arg-tRNA gene for SRX319065.14. One tRNA gene (Ile) was also predicted in reference virus – Pacmanvirus as described previously (Andreani et al., 2017). The complete statistics for the MAGs are provided in **Table 1**.

Phylogenetic Relationship Between the Asfarviruses

To assess the phylogenetic diversity and evolutionary relationships of the new MAGs, we constructed a phylogenetic tree based on alignment of the five conserved marker genes. These marker genes have been previously described to be highly conserved in the NCLDVs (Yutin et al., 2009; Moniruzzaman et al., 2020a). The phylogenetic analysis revealed that although the Asfarvirus MAGs formed clades with the five reference genomes (ASFV, Abalone asfarvirus, Kaumoebavirus, Faustovirus, and Pacmanvirus) in some cases, overall, the new MAGs had deep branches and were not closely related to reference viruses. The numerous deep-branching lineages in the tree underscores the high level of phylogenetic divergence between different Asfarviruses. The new MAGs were obtained from different environments, including freshwater (18), marine (12), landfill (2), non-marine saline lake (2), and mine tailing samples (1), highlighting their broad distribution. Clustering of the isolates according to the environment was also apparent in the phylogenetic tree, with several clades found only in marine or freshwater environments (**Figure 1**). This suggests that the broad habitat preference of many Asfarviruses may be conserved across some clades.

The MAG GVMAG-S-3300005056-23 was the most basal-branching Asfarvirus genome. We compared the proteins encoded in this genome to the NCBI RefSeq database and found that 13 had best hits to Poxviruses (compared to at most 4 in the other Asfarvirus MAGs), while 37 proteins had best hits to Asfarvirus genomes in this database (**Supplementary Data 1**). Together with its basal placement in our phylogeny, these results suggest that GVMAG-S-3300005056-23 is either a basal branching Asfarvirus or possibly even a member of a new family of NCLDV. We chose to use Poxviruses to root our phylogeny because this family is often considered to be most closely related to the *Asfarviridae* (Iyer et al., 2006; Koonin and Yutin, 2018), but it remains unclear where the root of the NCLDV should be placed, and other studies have recovered topologies that place the Asfarviruses as a sister group to other NCLDV families (Guglielmini et al., 2019). For purposes of our analysis, here, we kept GVMAG-S-3300005056-23 as a basal-branching Asfarvirus, but further studies are needed to confirm the evolutionary provenance of this MAG.

In addition to phylogenetic analysis, we also performed pairwise AAI analysis to assess the genomic divergence between different Asfarviruses. Our analysis recovered pairwise AAI values ranging from 27 to 75% (**Figure 2**) with mean and median values of 31.7 and 31.0%, respectively. This result is consistent

TABLE 1 | General statistics of the five Asfarvirus genomes and 35 viral MAGs.

MAGs	No. of contigs	Genome length	GC content	No. of proteins	N50 size	tRNA genes	Missing core genes
ERX552270.16	10	262,392	20.82	229	25,815	3	VLTF3
ERX556003.45	13	246,693	18.51	215	19,050	0	A32
GVMAG-M-3300000574-23	8	214,432	27.43	214	45,232	0	MCP
GVMAG-M-3300009068-46	18	220,122	25.09	188	13,641	0	–
GVMAG-M-3300009436-29	11	215,977	40.48	198	21,496	0	–
GVMAG-M-3300010160-26	6	221,296	29.40	199	55,434	0	–
GVMAG-M-3300013005-64	8	285,977	17.34	251	76,336	0	–
GVMAG-M-3300013133-40	15	171,139	34.27	158	11,819	1	MCP
GVMAG-M-3300014204-43	9	297,005	49.23	274	41,226	0	–
GVMAG-M-3300014204-45	10	202,564	37.00	189	20,298	0	–
GVMAG-M-3300017989-35	18	154,311	40.28	201	7,786	0	–
GVMAG-M-3300018416-36	19	208,202	25.28	188	13,575	0	–
GVMAG-M-3300020542-1	19	197,061	20.23	219	11,922	0	–
GVMAG-M-3300022309-7	19	310,111	20.07	325	23,777	0	PolB
GVMAG-M-3300022916-57	3	330,706	46.21	277	173,012	0	–
GVMAG-M-3300023174-150	21	225,080	29.51	220	13,647	0	MCP
GVMAG-M-3300023174-161	21	229,163	30.10	206	15,453	1	A32
GVMAG-M-3300023174-165	29	353,296	26.17	398	12,847	0	PolB
GVMAG-M-3300023184-110	14	474,957	25.61	372	44,583	0	–
GVMAG-M-3300023184-186	10	496,936	27.06	417	55,208	0	–
GVMAG-M-3300024062-1	14	349,792	21.91	377	56,296	0	–
GVMAG-M-3300027707-33	21	337,477	30.37	299	20,830	0	A32
GVMAG-M-3300027793-10	14	349,525	33.53	405	33,981	1	–
GVMAG-M-3300027833-19	11	297,297	30.60	324	73,615	0	–
GVMAG-S-1035124-107	7	239,202	18.99	200	55,079	0	A32
GVMAG-S-1092944-30	13	223,700	59.45	215	25,128	0	–
GVMAG-S-3300002466-141	10	331,811	39.41	420	49,921	0	–
GVMAG-S-3300005056-23	19	518,885	29.19	549	37,879	1	–
GVMAG-S-3300009702-144	28	580,795	28.00	591	38,326	0	–
GVMAG-S-3300010160-169	11	229,474	33.48	226	22,741	1	SFII
SRX319064.32	8	154,393	32.72	145	27,897	0	–
SRX319065.14	3	342,906	21.77	305	178,940	1	–
SRX327722.61	4	202,406	45.42	179	47,579	0	–
SRX802963.109	8	175,258	30.38	161	18,153	0	–
SRX802982.1	6	120,034	43.81	131	18,923	0	–
ASFV	–	170,101	38.95	–	–	0	–
Abalone asfarvirus	–	155,181	31.63	–	–	0	–
Faustovirus	–	466,265	36.22	–	–	0	–
Kaumoebavirus	–	350,731	43.7	–	–	0	–
Pacmanvirus	–	395,405	33.62	–	–	1	–

with the deep-branching clades identified in the phylogenetic analysis and confirms the high genomic divergence within the *Asfarviridae*.

Pan-Genomics of the Asfaviruses

We found 7,410 total OGs, including 6,480 that were found in one Asfarvirus genome only. The number of unique OGs for each genome ranged from 48 to 428. We observed 12 core OGs in 90% of genomes, including the MCP, VLTF3-like transcription factor, A32 packaging ATPase, DNA topoisomerase II, DNA ligase, DNA PolB, RNA polymerase subunit B, ATP dependent helicase hrpA, VVA8L-like transcription factor, and some hypothetical proteins (**Figure 3**). Nonetheless, the high number of genome-specific OGs highlights the genomic diversity present in the *Asfarviridae* family, which is consistent with the high level of variability in other families of NCLDV (Van Etten et al., 2010b).

In order to visualize the pattern of gene sharing, we performed bipartite network analysis using the Asfarvirus OGs, with six Poxvirus genomes used as non-Asfarvirus references. Given that virus evolution is characterized by extensive gene loss, gain, and exchange, this approach can be complementary to traditional phylogenetic analysis (Iranzo et al., 2016). The bipartite network showed some clustering of the MAGs based upon the habitat (**Figure 4**), although many co-clustered MAGs are also closely

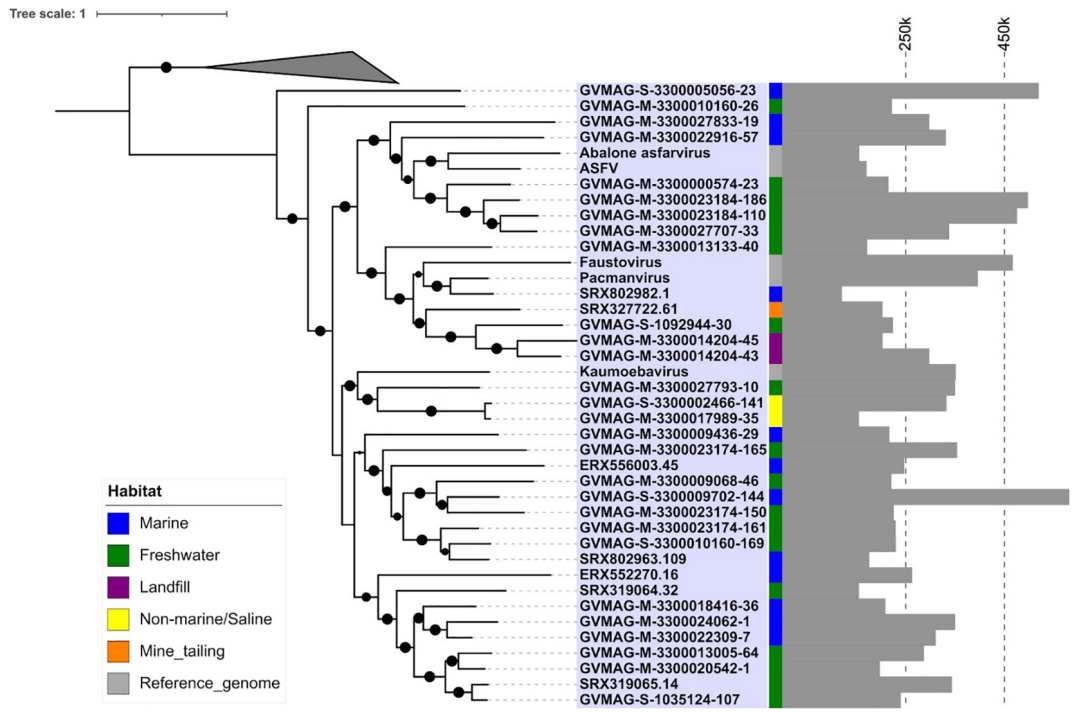


FIGURE 1 | Phylogenetic tree based on five conserved marker genes. (The inner strip represents the habitat while the bar chart with scale represents the genome size of the MAGs in bp.) The size of the black dot represents the bootstrap values. Only bootstrap values greater than 0.5 are shown.

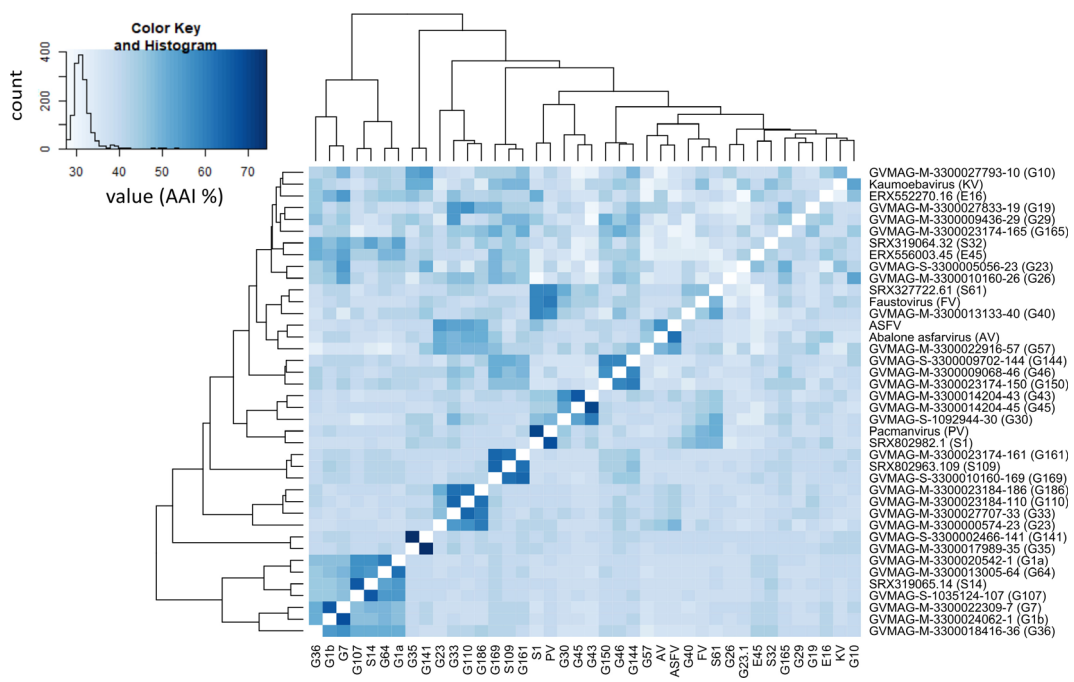


FIGURE 2 | Amino acid identity percentage between the MAGs and reference Asfarviruses. The histogram inside the color bar represents the frequency of AAI%.

related and common gene content due to shared ancestry cannot be ruled out. The *Poxviridae* clustered separately in a small sub-network, indicating that their gene content is clearly distinct from the *Asfarviridae*. Hence, the bipartite network provides support for the phylogenetic findings we have for the Asfaviruses and depicts the gene-sharing pattern of these viruses.

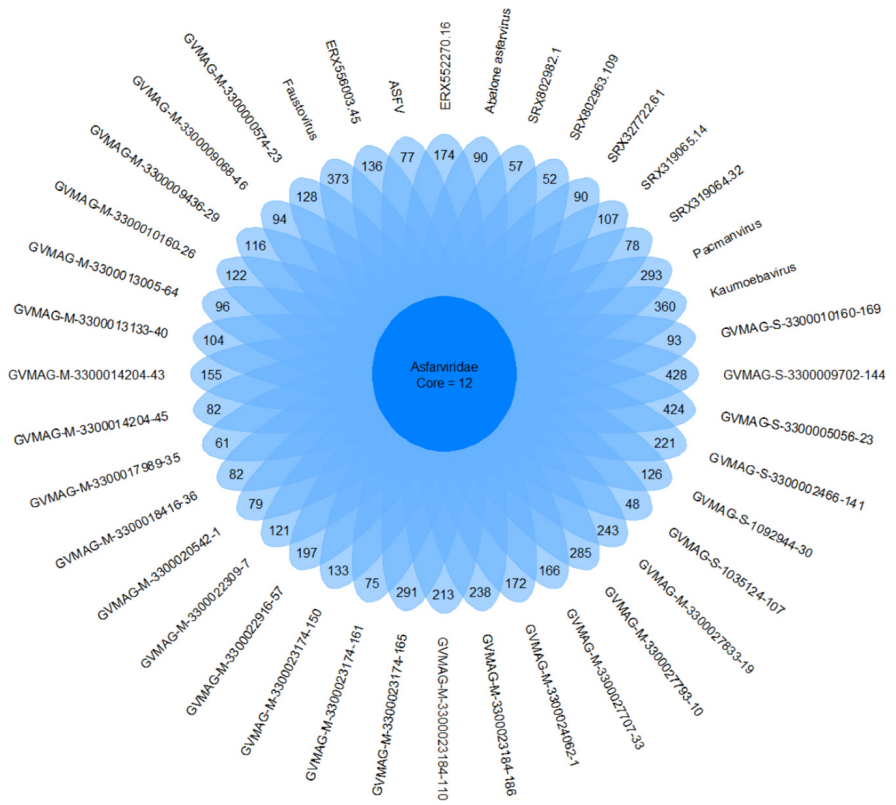


FIGURE 3 | Unique and core genes shared between the MAGs and reference Asfarviruses. Here, we define "core" as all genes found in 90% or more genomes.

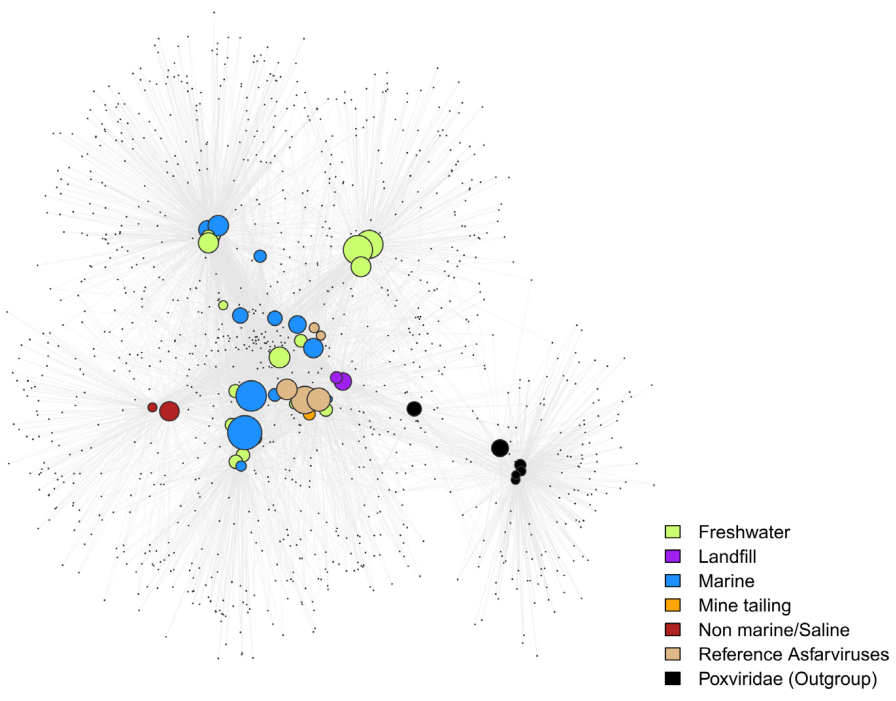


FIGURE 4 | Bipartite network plot for the MAGs. The larger nodes represent genomes while the smaller nodes represent OGs/gene families. Genomes were connected to the genes if they encode one. MAGs are colored based on their habitat. The size of the larger nodes represents genome size.

Genomic Chimerism of the Asfarviruses

Nucleo-cytoplasmic large DNA viruses are known to have chimeric genomes with genes that are derived from multiple sources (Boyer et al., 2009), and we therefore sought to quantify the extent of this genomic chimerism in environmental Asfarviruses by comparing the encoded proteins of the Asfarvirus MAGs to the RefSeq database (see section "Materials and Methods" for details; **Supplementary Data 1**). We found that between 40 and 70% of the proteins in each genome had no detectable hits to reference proteins, while 16–55% had best matches to other viruses, 5–22% to Eukaryotes, 3–15% to Bacteria, and 0–2% to Archaea (**Figure 5A**). We examined the proteins with best hits to Eukaryotes in more detail because this may provide some insight into host-virus gene exchange and therefore link these viruses to putative hosts. Overall, best hits to eukaryotes included matches to Animalia, Plantae, Fungi, and Protists such as Stramenopiles, Alveolata, Archaeplastida, Cryptista, Excavata, Choanomonada, Apusozoa, Porifera, and Amoebozoa (**Figure 5B**). The percent identity of these matches ranged from 19.4 to 93.2 (median 35.3), with only 4 greater than 90%, suggesting that, if these represent gene exchanges between NCLDV and eukaryotes, the vast majority have not occurred recently. Although recent studies have revealed a dynamic gene exchange between NCLDV and eukaryotic lineages that can be used to link viruses to their hosts (Moniruzzaman et al., 2020b; Schulz et al., 2020), our analysis did not identify any clear signatures in the Asfarvirus MAGs that could be used for this purpose. It is possible that future work examining endogenous NCLDV signatures in eukaryotic genomes may be useful to better identify virus-host relationships.

Asfarvirus Genes Involved in Manipulating Host Metabolism

To assess the potential functions of the proteins encoded by the MAGs, we performed functional annotation using HMMER searches against the EggNOG database (all annotations available in **Supplementary Data 2**). As expected, in all MAGs we detected genes involved in DNA replication and repair, transcription, and post-translational modification, which is consistent with the prevalence of these functions across NCLDV (Yutin and Koonin, 2012; **Figure 6**). Among the proteins involved in post-translational modification, we found genes responsible for ubiquitination (KOG0802 and KOG1812) and ubiquitin dependent proteins in 26 MAGs. Ubiquitination has been found to be an important counteracting mechanism to oxidative stress response in eukaryotes that direct the unwanted proteins to proteasome for degradation (Silva et al., 2015). In *Aureococcus anophagefferens* giant virus (AaV), ubiquitin dependent protein-ubiquitin ligases were found to be expressed within 5 min of virus-infection and is thought to be involved in degradation of host proteins (Moniruzzaman et al., 2018). The ubiquitin protein has also been reported in Marseilleviruses, where it is thought to play an important role in host signaling (Boyer et al., 2009). A protein homologous to the ubiquitin-proteasome (UP) system has been found to be encoded by ASFV, suggesting its role during early infection and replication (Barrado-Gil et al., 2017). Hence, this suggests that ubiquitination may be a common mechanism across diverse Asfarviruses.

Genes predicted to be involved in carbohydrate metabolism were prevalent in the MAGs, consistent with previous findings that these genes are widespread in NCLDVs. We observed glycosyltransferase enzymes that are important in glycosylation of viral proteins in 15 Asfarvirus MAGs. These enzymes have been previously reported in giant viruses (Markine-Goriaynoff et al., 2004). Also, past studies have indicated the presence of glycosylating genes (Van Etten et al., 2010a; Piacente et al., 2015) and other enzymes involved in carbohydrate metabolism in NCLDVs (Fischer et al., 2010). Interestingly, we found genes involved in the shikimate pathway that is linked to the biosynthesis and metabolism of carbohydrates and aromatic amino acids (phenylalanine, tryptophan, and tyrosine) in five MAGs. We found 3-deoxy-7-phosphoheptulonate synthase (2QPSU) (the first enzyme in the shikimate pathway), chorismate synthase (KOG4492), and prephenate dehydrogenase (KOG2380) all in ERX556003.45 and only 3-deoxy-7-phosphoheptulonate synthase in four other MAGs. The shikimate pathway is widespread in bacteria, archaea, and protists but not in metazoans (Richards et al., 2006). We also found acetolactate synthase genes (KOG4166) in three MAGs. Acetolactate synthase that are involved in the synthesis of amino acids such as leucine, isoleucine, and valine has been previously described to be present in large DNA viruses infecting green algae mainly, *Prasinovirus* (Weynberg et al., 2009; Moreau et al., 2010; Zhang et al., 2015). Hence, the detection of these enzymes shows the potential role of the Asfarvirus MAGs in the manipulation of amino acid metabolism in their hosts during infection.

Genes responsible for signal transduction were also present in some of the MAGs. KOGs representing serine/threonine protein kinase and tyrosine/serine/threonine phosphatase were present in seven MAGS. These enzymes constitute a major form of signaling and regulation of many cellular pathways such as cell proliferation, differentiation, and cell death. Serine/threonine kinases have also been reported in Marseillevirus, Iridovirus, and Ascoviruses (Boyer et al., 2009; Piégu et al., 2015) and ASFV, suggesting that it might have a role in early infection and programmed cell death (apoptosis) (Baylis et al., 1993).

We found genes homologous to cysteine desulfurase (COG1104) proteins in 21 out of 35 MAGs (**Supplementary Data 2**). NifS genes whose presumed functions are similar to that of cysteine desulfurase are reported to be associated with ASFV, Faustovirus, and Pacmanvirus with possible involvement in host cell interactions (Andreani et al., 2017). Cysteine desulfurase proteins are found in bacteria and eukaryotes and are involved in the biosynthesis of iron (Fe) – sulphur (S) clusters, thiamine, biotin, lipoic acid, molybdopterin, NAD, and thionucleosides in tRNA (Mihara and Esaki, 2002). Hence, the discovery of the enzyme cysteine desulfurase adds to the viral proteins involved in electron transfer processes.

Gene encoding for cell redox homeostasis (KOG0191) and cellular response to nitrogen starvation (KOG1654) were also common among the MAGs. Nutrient limitation has the potential

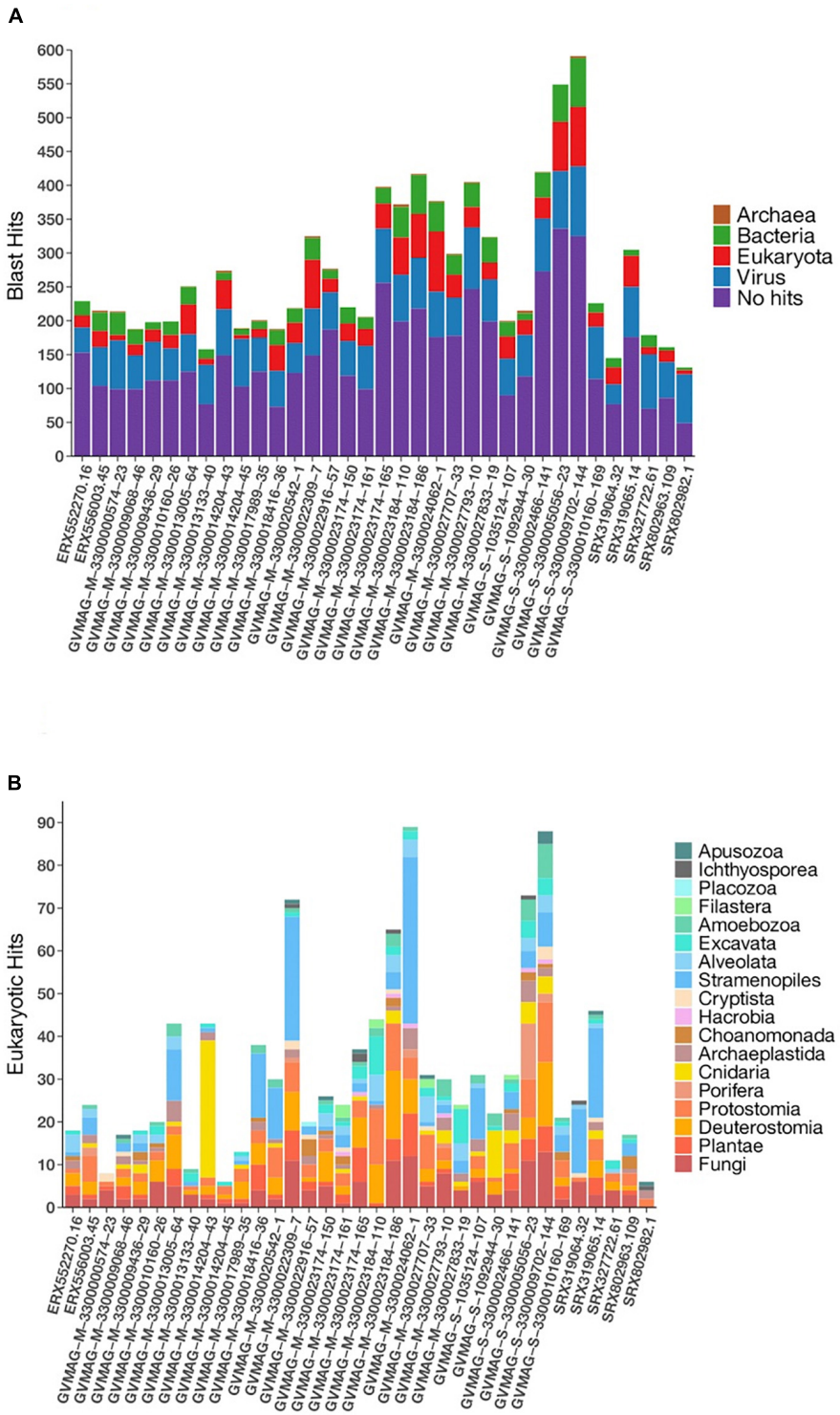


FIGURE 5 | Distribution of homologous hits to MAGs determined by the BLASTp. **(A)** Total hits to three domains of life and viruses **(B)** eukaryotic hits.

to reduce viral productivity; virus reproduction mostly depends upon the intracellular nitrogen and phosphorous pool during early infection while they might depend upon the extracellular nitrogen availability as infection proceeds (Zimmerman et al., 2020). Genes involved in responding to nutrient starvation can influence the nutrient uptake and replication in these viruses. Overall, these results demonstrate that in addition to universal genes that play a role in host invasion and viral replication,

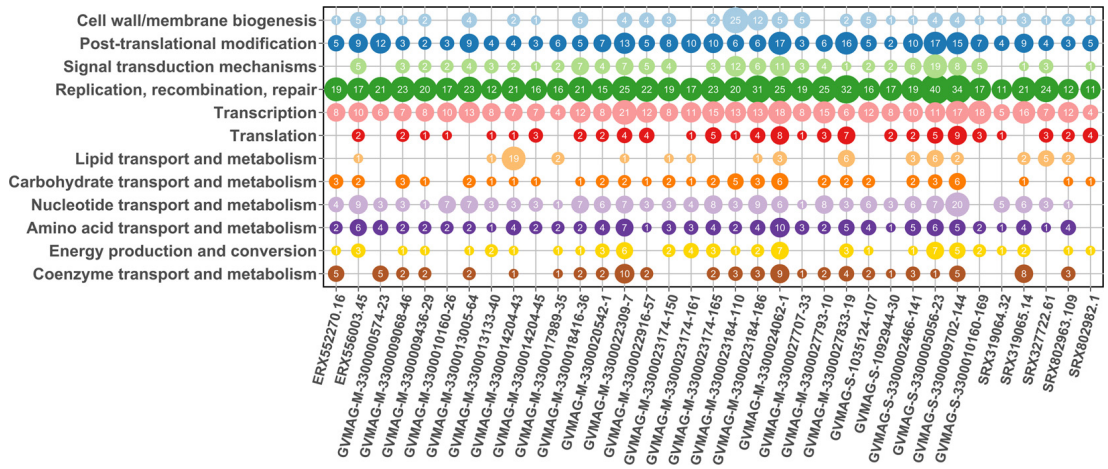


FIGURE 6 | Protein annotation for MAGs. The *x*-axis represents the MAGs while *y*-axis represents the COG category. The number inside the bubble represents the number of genes present in that MAG that had the annotated function.

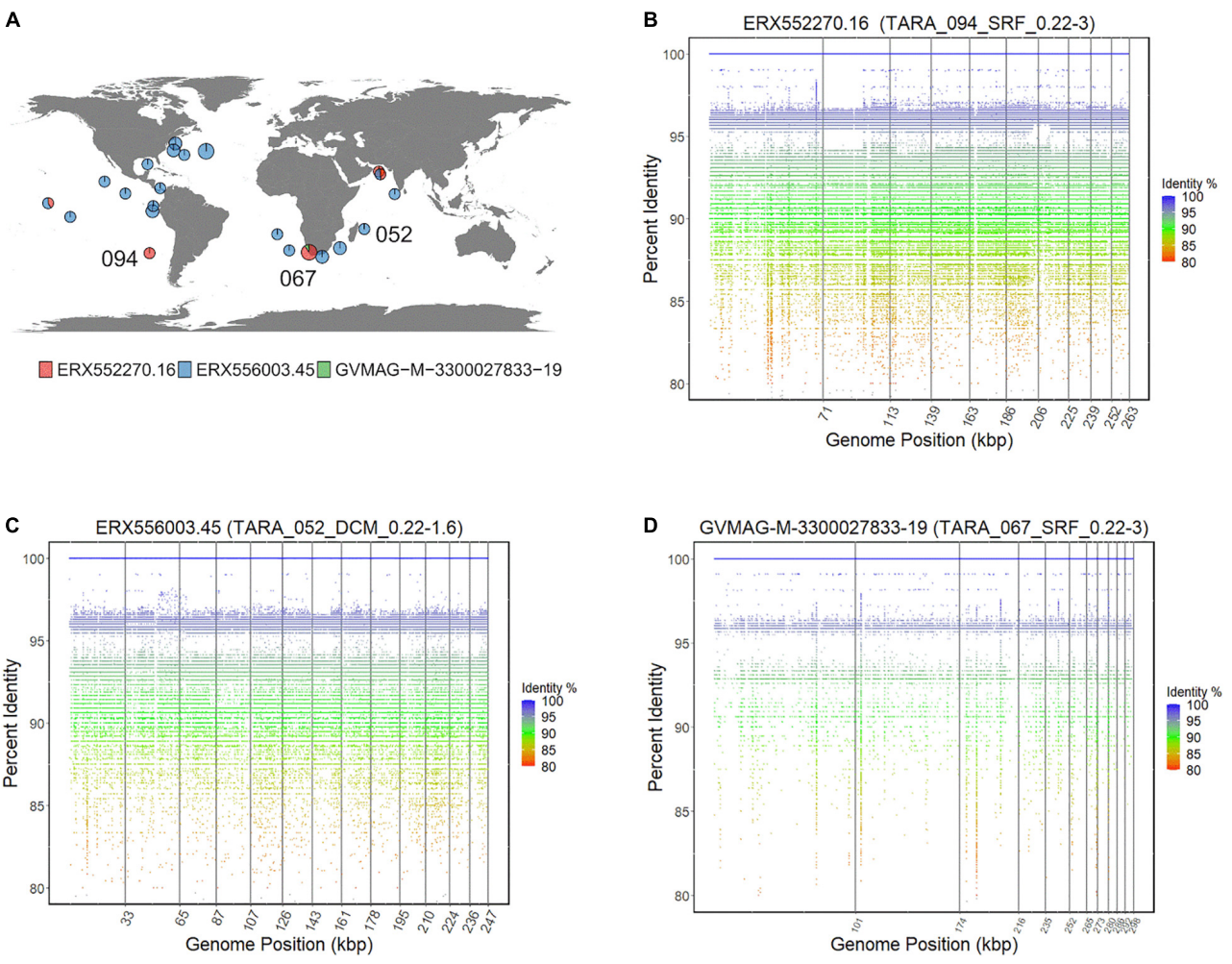


FIGURE 7 | (A) Distribution of Asfarvirus matching metagenomic reads from the TARA ocean project. **(B–D)** Fragment recruitment plot for metagenomic reads to ERX552270.16, ERX556003.45, and GVMAG-M-3300027833-19, respectively. The *x*-axis of the recruitment plot shows position of the metagenomic reads along the genome length and *y*-axis represents the percent identity.

TABLE 2 | Amino acid identity between the HcDNAV genes (only genes available at NCBI) and MAG ERX552270.16 as analyzed by blastp.

HcDNAV genes	MAG genes	AAI %
HNH endonuclease (YP_009507839.1)	contig_19878_13	95.88
DNA mismatch repair protein (BAJ49801.1)	contig_7191_19	98
Type B DNA polymerase (YP_009507841.1)	contig_19878_12	98.37
DNA-directed RNA polymerase subunit (BAI48199.1)	contig_19878_28	98.97

Asfarviruses also contain genes involved in metabolism, hence, capable of reprogramming cells into virocells during infection (Moniruzzaman et al., 2020a).

Biogeography of Marine Asfarviruses

While ASFV is a terrestrial pathogen and most cultured Asfarviruses were isolated from sewage samples, various metagenomic studies have revealed that NCLDVs are highly diverse and abundant in aquatic environments (Monier et al., 2008; Hingamp et al., 2013), and one recent study noted that Asfarviruses are prevalent in some marine samples (Endo et al., 2020). To examine the biogeography of the Asfarvirus MAGs in more detail we conducted a fragment recruitment analysis using reads from the Tara oceans expedition (Sunagawa et al., 2015). We examined 28 diverse metagenomic samples from surface and deep chlorophyll maxima (DCM) oceanic regions. The Asfarvirus MAG ERX552270.16 was present in eight metagenomic samples (from five different TARA stations), ERX556003.45 was found in 19, and GVMAG-M-3300027833-19 was found in one, revealing that some Asfarvirus are globally distributed in the ocean (**Figure 7A**). The fragment recruitment plots revealed that the MAGs had consistent coverage of reads with 100% nucleic acid identity matches to the metagenomic reads (**Figures 7B–D** and **Supplementary Figures 1, 2**), demonstrating high similarity of these viruses across long distances. Few gaps were visible in the recruitment plots, indicating the absence of readily-identifiable genomic islands in these viruses.

Previous studies have shown that the virus HcDNAV infects the marine dinoflagellate *Heterocapsa circularisquama*, which is responsible for harmful algal blooms in the marine environment (Tarutani et al., 2001; Nagasaki et al., 2003). This is notable since very few viruses that infect dinoflagellates have been characterized, and of these HcDNAV is the only large DNA virus (Nagasaki, 2008). Although a complete genome of HcDNAV is not available, several marker genes from this virus have been sequenced, are available in NCBI and have been previously reported (Ogata et al., 2009). We found that the MAG ERX552270.16 bore high sequence similarity to the HcDNAV marker genes, indicating that this MAG represents a closely related virus that potentially infects the same host. The Family B Polymerase (YP_009507841.1), HNH endonuclease (YP_009507839.1), DNA directed RNA Polymerase (BAI48199.1), and DNA mismatch repair protein (mutS) (BAJ49801.1) of HcDNAV all had 95.8 to 99% AAI to homologs in ERX552270.16 (**Table 2**). The PolB enzyme of

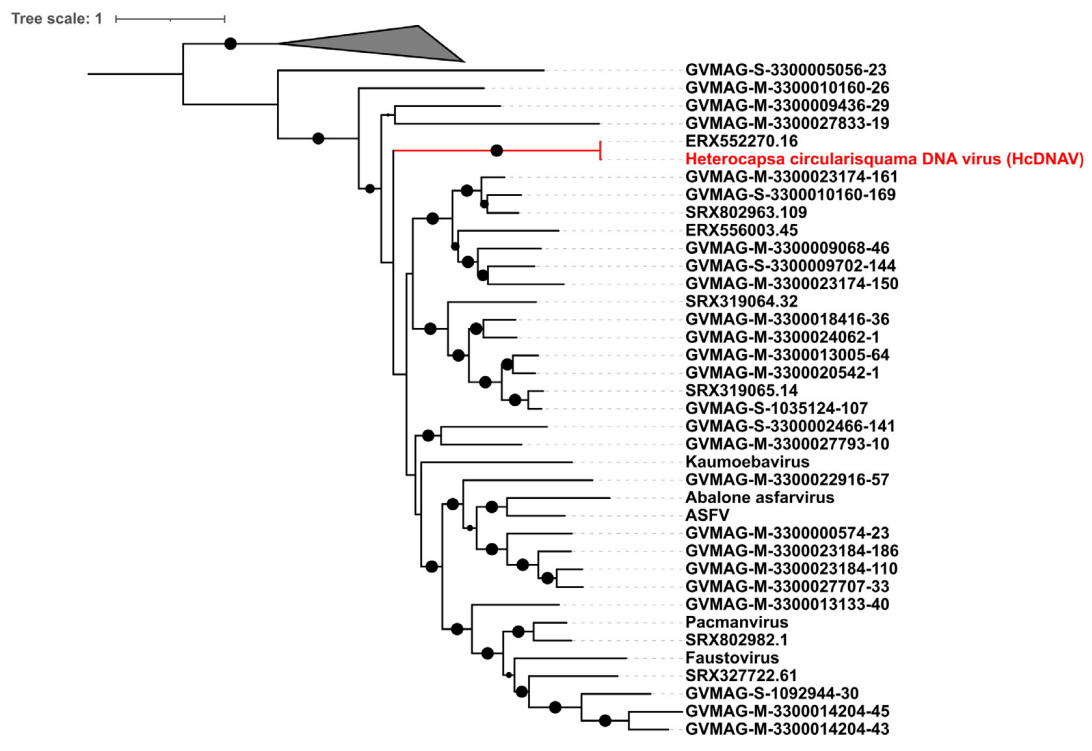


FIGURE 8 | Phylogenetic tree reconstruction based on DNA polymerase B gene (New reference virus HcDNAV has been added). The size of the black dot represents the bootstrap values. Only bootstrap values greater than 0.5 are shown.

ERX552270.16 also contained the notable YSDTDS motif that was previously found in HcDNAV (Ogata et al., 2009). Moreover, we constructed a PolB phylogeny of the *Asfarviridae* that confirmed that these viruses cluster closely together (**Figure 8**). Our fragment recruitment analysis from Tara Oceans data confirmed that ERX552270.16 is widespread in the ocean, especially in costal environments (**Figure 7** and **Supplementary Figure 1**), consistent with the hypothesis that it is a marine virus that also infects *Heterocapsa circularisquama* or a closely related dinoflagellate. Given these similarities, ERX552270.16 may be a useful reference genome for exploring the genomics and distribution of close relatives of HcDNAV, though further work will be necessary to confirm the host of ERX552270.16.

CONCLUSION

While ASFV was the only known member of *Asfarviridae* for many years, recent work has identified numerous additional members of this viral family. In this study, we provide a robust phylogenetic and comparative genomic analysis of this viral family. Our results highlight the high level of genomic and phylogenetic divergence between disparate members of the *Asfarviridae*, and homology searches suggest that many genes within this viral group are potentially the product of ancient horizontal transfers from cellular lineages. Moreover, we provide fragment recruitment plots that confirm that some Asfarviruses are ubiquitous in the ocean, where they may infect ecologically important protists such as bloom forming dinoflagellates. These findings suggest that diverse Asfarviruses are broadly distributed in the environment and play important roles in numerous ecosystems.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found here: https://www.ncbi.nlm.nih.gov/.

AUTHOR CONTRIBUTIONS

FA designed the study. SK and MM performed the experiment. SK and FA wrote the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

This research was funded by a Simons Foundation Early Career Award in Marine Microbial Ecology and Evolution and an NSF IIBR award 1918271 to FA.

ACKNOWLEDGMENTS

We acknowledge the use of the Virginia Tech Advanced Research Computing Center for bioinformatic analyses performed in this study. We are thankful to the members of Aylward Lab for their helpful suggestions.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2021.657471/full#supplementary-material

REFERENCES

Abergel, C., Legendre, M., and Claverie, J.-M. (2015). The rapidly expanding universe of giant viruses: Mimivirus, Pandoravirus, Pithovirus and Mollivirus. *FEMS Microbiol. Rev.* 39, 779–796. doi: 10.1093/femsre/fuv037

Andreani, J., Khalil, J. Y. B., Sevvana, M., Benamar, S., Di Pinto, F., Bitam, I., et al. (2017). Pacmanvirus, a new giant icosahedral virus at the crossroads between Asfarviridae and Faustoviruses. *J. Virol.* 91, e212–17. doi: 10.1128/JVI.00212-17

Aylward, F. O., and Moniruzzaman, M. (2021). ViralRecall: a flexible command-line tool for the detection of giant virus signatures in omic data. *Viruses* 13:150. doi: 10.3390/v13020150

Bäckström, D., Yutin, N., Jørgensen, S. L., Dharamshi, J., Homa, F., Zaremba-Niedwiedzka, K., et al. (2019). Virus genomes from deep sea sediments expand the ocean megavirome and support independent origins of viral gigantism. *MBio* 10, e2497–18. doi: 10.1128/mBio.02497-18

Bajrai, L. H., Benamar, S., Azhar, E. I., Robert, C., Levasseur, A., Raoult, D., et al. (2016). Kaumoebavirus, a new virus that clusters with faustoviruses and Asfarviridae. *Viruses* 8:278. doi: 10.3390/v8110278

Barrado-Gil, L., Galindo, I., Martínez-Alonso, D., Viedma, S., and Alonso, C. (2017). The ubiquitin-proteasome system is required for African swine fever replication. *PLoS One* 12:e0189741. doi: 10.1371/journal.pone.0189741

Baylis, S. A., Banham, A. H., Vydelingum, S., Dixon, L. K., and Smith, G. L. (1993). African swine fever virus encodes a serine protein kinase which is packaged into virions. *J. Virol.* 67, 4549–4556. doi: 10.1128/jvi.67.8.4549-4556.1993

Benamar, S., Reteno, D. G. I., Bandaly, V., Labas, N., Raoult, D., and La Scola, B. (2016). Faustoviruses: comparative genomics of new megavirales family members. *Front. Microbiol.* 7:3. doi: 10.3389/fmicb.2016.00003

Boyer, M., Yutin, N., Pagnier, I., Barrassi, L., Fournous, G., Espinosa, L., et al. (2009). Giant Marseillevirus highlights the role of amoebae as a melting pot in emergence of chimeric microorganisms. *Proc. Natl. Acad. Sci. U.S.A.* 106, 21848–21853. doi: 10.1073/pnas.0911354106

Capella-Gutierrez, S., Silla-Martinez, J. M., and Gabaldon, T. (2009). trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* 25, 1972–1973. doi: 10.1093/bioinformatics/btp348

Csardi, G., and Nepusz, T. (2006). The igraph software package for complex network research. *Int. J. Complex Syst.* 1695, 1–9.

Eddy, S. R. (2011). Accelerated profile HMM searches. *PLoS Comput. Biol.* 7:e1002195. doi: 10.1371/journal.pcbi.1002195

Endo, H., Blanc-Mathieu, R., Li, Y., Salazar, G., Henry, N., Labadie, K., et al. (2020). Biogeography of marine giant viruses reveals their interplay with eukaryotes and ecological functions. *Nat. Ecol. Evol.* 4, 1639–1649. doi: 10.1038/s41559-020-01288-w

Fischer, M. G., Allen, M. J., Wilson, W. H., and Suttle, C. A. (2010). Giant virus with a remarkable complement of genes infects marine zooplankton. *Proc. Natl Acad. Sci. U.S.A.* 107, 19508–19513. doi: 10.1073/pnas.1007615107

Guglielmini, J., Woo, A., Krupovic, M., Forterre, P., and Gaia, M. (2019). Diversification of giant and large eukaryotic dsDNA viruses predated the origin of modern eukaryotes. *Proc. Natl. Acad. Sci. U.S.A.* 116, 19585–19592. doi: 10.1101/455816

Hingamp, P., Grimsley, N., Acinas, S. G., Clerissi, C., Subirana, L., Poulain, J., et al. (2013). Exploring nucleo-cytoplasmic large DNA viruses in Tara Oceans microbial metagenomes. *ISME J.* 7, 1678–1695. doi: 10.1038/ismej.2013.59

Hoang, D. T., Chernomor, O., von Haeseler, A., Minh, B. Q., and Vinh, L. S. (2018). UFBoot2: improving the ultrafast bootstrap approximation. *Mol. Biol. Evol.* 35, 518–522. doi: 10.1093/molbev/msx281

Huerta-Cepas, J., Szklarczyk, D., Forslund, K., Cook, H., Heller, D., Walter, M. C., et al. (2016). eggNOG 4.5: a hierarchical orthology framework with improved functional annotations for eukaryotic, prokaryotic and viral sequences. *Nucleic Acids Res.* 44, D286–D293.

Hyatt, D., Chen, G.-L., Locascio, P. F., Land, M. L., Larimer, F. W., and Hauser, L. J. (2010). Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* 11:119. doi: 10.1186/1471-2105-11-119

Iranzo, J., Krupovic, M., and Koonin, E. V. (2016). The double-stranded DNA virosphere as a modular hierarchical network of gene sharing. *MBio* 7, e978–16. doi: 10.1128/mBio.00978-16

Iyer, L. M., Balaji, S., Koonin, E. V., and Aravind, L. (2006). Evolutionary genomics of nucleo-cytoplasmic large DNA viruses. *Virus Res.* 117, 156–184. doi: 10.1016/j.virusres.2006.01.009

Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K. F., von Haeseler, A., and Jermiin, L. S. (2017). ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat. Methods* 14, 587–589. doi: 10.1038/nmeth.4285

Kiełbasa, S. M., Wan, R., Sato, K., Horton, P., and Frith, M. C. (2011). Adaptive seeds tame genomic sequence comparison. *Genome Res.* 21, 487–493. doi: 10.1101/gr.113985.110

Koonin, E. V., Dolja, V. V., Krupovic, M., Varsani, A., Wolf, Y. I., Yutin, N., et al. (2020). Global organization and proposed megataxonomy of the virus world. *Microbiol. Mol. Biol. Rev.* 84, e0061–19. doi: 10.1128/MMBR.00061-19

Koonin, E. V., and Yutin, N. (2018). Multiple evolutionary origins of giant viruses [version 1; peer review: 4 approved]. *F1000Research* 7:1840. doi: 10.12688/f1000research.16248.1

Laslett, D. (2004). ARAGORN, a program to detect tRNA genes and tmRNA genes in nucleotide sequences. *Nucleic Acids Res.* 32, 11–16. doi: 10.1093/nar/gkh152

Lechner, M., Findeiß, S., Steiner, L., Marz, M., Stadler, P. F., and Prohaska, S. J. (2011). Proteinortho: Detection of (Co-)orthologs in large-scale analysis. *BMC Bioinformatics* 12:124. doi: 10.1186/1471-2105-12-124

Letunic, I., and Bork, P. (2019). Interactive Tree Of Life (iTOL) v4: recent updates and new developments. *Nucleic Acids Res.* 47, W256–W259.

Markine-Goriaynoff, N., Gillet, L., Van Etten, J. L., Korres, H., Verma, N., and Vanderplasschen, A. (2004). Glycosyltransferases encoded by viruses. *J. Gen. Virol.* 85, 2741–2754. doi: 10.1099/vir.0.80320-0

Matsuyama, T., Takano, T., Nishiki, I., Fujiwara, A., Kiryu, I., Inada, M., et al. (2020). A novel Asfarvirus-like virus identified as a potential cause of mass mortality of abalone. *Sci. Rep.* 10:4620.

Mihara, H., and Esaki, N. (2002). Bacterial cysteine desulfurases: their function and mechanisms. *Appl. Microbiol. Biotechnol.* 60, 12–23. doi: 10.1007/s00253-002-1107-4

Minh, B. Q., Schmidt, H. A., Chernomor, O., Schrempf, D., Woodhams, M. D., von Haeseler, A., et al. (2020). IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Mol. Biol. Evol.* 37, 1530–1534. doi: 10.1093/molbev/msaa015

Monier, A., Claverie, J.-M., and Ogata, H. (2008). Taxonomic distribution of large DNA viruses in the sea. *Genome Biol.* 9:R106.

Moniruzzaman, M., Gann, E. R., and Wilhelm, S. W. (2018). Infection by a Giant Virus (AaV) induces widespread physiological reprogramming in CCMP1984 a harmful bloom algae. *Front. Microbiol.* 9:752. doi: 10.3389/fmicb.2018.00752

Moniruzzaman, M., Martinez-Gutierrez, C. A., Weinheimer, A. R., and Aylward, F. O. (2020a). Dynamic genome evolution and complex virocell metabolism of globally-distributed giant viruses. *Nat. Commun.* 11:1710.

Moniruzzaman, M., Weinheimer, A. R., Martinez-Gutierrez, C. A., and Aylward, F. O. (2020b). Widespread endogenization of giant viruses shapes genomes of green algae. *Nature* 588, 141–145. doi: 10.1038/s41586-020-2924-2

Montgomery, R. E., and Eustace Montgomery, R. (1921). On a form of swine fever occurring in British East Africa (Kenya Colony). *J. Comp. Pathol. Ther.* 34, 159–191. doi: 10.1016/s0368-1742(21)80031-4

Moreau, H., Piganeau, G., Desdevises, Y., Cooke, R., Derelle, E., and Grimsley, N. (2010). Marine prasinovirus genomes show low evolutionary divergence and acquisition of protein metabolism genes by horizontal gene transfer. *J. Virol.* 84, 12555–12563. doi: 10.1128/jvi.01123-10

Nagasaki, K. (2008). Dinoflagellates, diatoms, and their viruses. *J. Microbiol.* 46, 235–243. doi: 10.1007/s12275-008-0098-y

Nagasaki, K., Tomaru, Y., Tarutani, K., Katanozaka, N., Yamanaka, S., Tanabe, H., et al. (2003). Growth characteristics and intraspecies host specificity of a large virus infecting the dinoflagellate Heterocapsa circularisquama. *Appl. Environ. Microbiol.* 69, 2580–2586. doi: 10.1128/aem.69.5.2580-2586.2003

Ogata, H., Toyoda, K., Tomaru, Y., Nakayama, N., Shirai, Y., Claverie, J.-M., et al. (2009). Remarkable sequence similarity between the dinoflagellate-infecting marine girus and the terrestrial pathogen African swine fever virus. *Virol. J.* 6:178. doi: 10.1186/1743-422x-6-178

O'Leary, N. A., Wright, M. W., Rodney Brister, J., Ciufo, S., Haddad, D., McVeigh, R., et al. (2016). Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res.* 44, D733–D745. doi: 10.1093/nar/gkv1189

Philippe, N., Legendre, M., Doutre, G., Couté, Y., Poirot, O., Lescot, M., et al. (2013). Pandoraviruses: amoeba viruses with genomes up to 2.5 Mb reaching that of parasitic eukaryotes. *Science* 341, 281–286. doi: 10.1126/science.1239181

Piacente, F., Gaglianone, M., Laugieri, M. E., and Tonetti, M. G. (2015). The autonomous glycosylation of large DNA Viruses. *Int. J. Mol. Sci.* 16, 29315–29328. doi: 10.3390/ijms161226169

Piégu, B., Asgari, S., Bideshi, D., Federici, B. A., and Bigot, Y. (2015). Evolutionary relationships of iridoviruses and divergence of ascoviruses from invertebrate iridoviruses in the superfamily Megavirales. *Mol. Phylogenet. Evol.* 84, 44–52. doi: 10.1016/j.ympev.2014.12.013

Raoult, D., Audic, S., Robert, C., Abergel, C., Renesto, P., Ogata, H., et al. (2004). The 1.2-megabase genome sequence of Mimivirus. *Science* 306, 1344–1350. doi: 10.1126/science.1101485

Reteno, D. G., Benamar, S., Khalil, J. B., Andreani, J., Armstrong, N., Klose, T., et al. (2015). Faustovirus, an asfarvirus-related new lineage of giant viruses infecting amoebae. *J. Virol.* 89, 6585–6594.

Richards, T. A., Dacks, J. B., Campbell, S. A., Blanchard, J. L., Foster, P. G., McLeod, R., et al. (2006). Evolutionary origins of the eukaryotic shikimate pathway: gene fusions, horizontal gene transfer, and endosymbiotic replacements. *Eukaryot. Cell* 5, 1517–1531. doi: 10.1128/ec.00106-06

Schulz, F., Roux, S., Paez-Espino, D., Jungbluth, S., Walsh, D. A., Denef, V. J., et al. (2020). Giant virus diversity and host interactions through global metagenomics. *Nature* 578, 432–436. doi: 10.1038/s41586-020-1957-x

Schvarcz, C. R., and Steward, G. F. (2018). A giant virus infecting green algae encodes key fermentation genes. *Virology* 518, 423–433. doi: 10.1016/j.virol.2018.03.010

Shen, W., Le, S., Li, Y., and Hu, F. (2016). SeqKit: a cross-platform and ultrafast toolkit for FASTA/Q File Manipulation. *PLoS One* 11:e0163962. doi: 10.1371/journal.pone.0163962

Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., et al. (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* 7:539. doi: 10.1038/msb.2011.75

Silva, G. M., Finley, D., and Vogel, C. (2015). K63 polyubiquitination is a new modulator of the oxidative stress response. *Nat. Struct. Mol. Biol.* 22, 116–123. doi: 10.1038/nsmb.2955

Sunagawa, S., Coelho, L. P., Chaffron, S., Kultima, J. R., Labadie, K., Salazar, G., et al. (2015). Ocean plankton. Structure and function of the global ocean microbiome. *Science* 348:1261359.

Tarutani, K., Nagasaki, K., Itakura, S., and Yamaguchi, M. (2001). Isolation of a virus infecting the novel shellfish-killing dinoflagellate *Heterocapsa circularisquama*. *Aquatic Microb. Ecol.* 23, 103–111. doi: 10.3354/ame023103

Van Etten, J. L., Gurnon, J. R., Yanai-Balser, G. M., Dunigan, D. D., and Graves, M. V. (2010a). Chlorella viruses encode most, if not all, of the machinery to glycosylate their glycoproteins independent of the endoplasmic reticulum and Golgi. *Biochimica Biophys. Acta (BBA) Gen. Subjects* 1800, 152–159. doi: 10.1016/j.bbagen.2009.07.024

Van Etten, J. L., Lane, L. C., and Dunigan, D. D. (2010b). DNA Viruses: the really big ones (Giruses). *Ann. Rev. Microbiol.* 64, 83–99. doi: 10.1146/annurev.micro.112408.134338

Warnes, G. R., Bolker, B., Bonebakker, L., Gentleman, R., Huber, W., Liaw, A., et al. (2020). *gplots: Various R Programming Tools for Plotting Data. R package*

version Q16 3.1.0. Available online at: https://cran.r-project.org/package=gplots (accessed November 9, 2020).

Weynberg, K. D., Allen, M. J., Ashelford, K., Scanlan, D. J., and Wilson, W. H. (2009). From small hosts come big viruses: the complete genome of a secondOstreococcus taurivirus, OtV-1. *Environ. Microbiol.* 11, 2821–2839. doi: 10.1111/j.1462-2920.2009.01991.x

Wickham, H. (2009). *ggplot2: Elegant Graphics for Data Analysis*. Berlin: Springer Science & Business Media.

Yutin, N., and Koonin, E. V. (2012). Hidden evolutionary complexity of Nucleo-Cytoplasmic Large DNA viruses of eukaryotes. *Virol. J.* 9:161.

Yutin, N., Wolf, Y. I., Raoult, D., and Koonin, E. V. (2009). Eukaryotic large nucleo-cytoplasmic DNA viruses: clusters of orthologous genes and reconstruction of viral genome evolution. *Virol. J.* 6:223. doi: 10.1186/1743-422x-6-223

Zhang, W., Zhou, J., Liu, T., Yu, Y., Pan, Y., Yan, S., et al. (2015). Four novel algal virus genomes discovered from Yellowstone Lake metagenomes. *Sci. Rep.* 5:15131.

Zimmerman, A. E., Howard-Varona, C., Needham, D. M., John, S. G., Worden, A. Z., Sullivan, M. B., et al. (2020). Metabolic and biogeochemical consequences of viral infection in aquatic ecosystems. *Nat. Rev. Microbiol.* 18, 21–34. doi: 10.1038/s41579-019-0270-x

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Microbiology

ORIGINAL RESEARCH
published: 08 April 2021
doi: 10.3389/fmicb.2021.663506

A Conserved Phenylalanine Residue of Autographa Californica Multiple Nucleopolyhedrovirus AC75 Protein Is Required for Occlusion Body Formation

Xingang Chen[1,2], Jian Yang[1,2], Xiaoqin Yang[1,2], Chengfeng Lei[1], Xiulian Sun[1] and Jia Hu[1]**

[1] *Wuhan Institute of Virology, Center for Biosafety Mega-Science, Chinese Academy of Sciences, Wuhan, China,* [2] *College of Life Sciences, University of Chinese Academy of Sciences, Beijing, China*

OPEN ACCESS

Edited by:
Julien Andreani,
IHU Mediterranee Infection, France

Reviewed by:
Peter Krell,
University of Guelph, Canada
Yves Gaudin,
Centre National de la Recherche Scientifique (CNRS), France

***Correspondence:**
Xiulian Sun
sunxl@wh.iov.cn
Jia Hu
hujia@wh.iov.cn

Specialty section:
This article was submitted to Virology, a section of the journal Frontiers in Microbiology

Received: *03 February 2021*
Accepted: *16 March 2021*
Published: *08 April 2021*

Citation:
Chen X, Yang J, Yang X, Lei C, Sun X and Hu J (2021) A Conserved Phenylalanine Residue of Autographa Californica Multiple Nucleopolyhedrovirus AC75 Protein Is Required for Occlusion Body Formation. Front. Microbiol. 12:663506. doi: 10.3389/fmicb.2021.663506

Autographa californica multiple nucleopolyhedrovirus (AcMNPV) *orf75* (*ac75*) is a highly conserved gene that is essential for AcMNPV propagation. However, the key domains or residues of the AC75 protein that play a role in viral propagation have not been identified. In this study, sequence alignment revealed that residues Phe-54 and Gln-81 of AC75 were highly conserved among alphabaculoviruses and betabaculoviurses. Thus, Phe-54 and Gln-81 AC75 mutation bacmids were constructed. We found that Gln-81 was not required for viral propagation, whereas mutating Phe-54 reduced budded virus production by 10-fold and impaired occlusion body formation when compared with that of the wild-type AcMNPV. Electron microscopy observations showed that the Phe-54 mutation affected polyhedrin assembly and also occlusion-derived virus embedding, whereas western blot analysis revealed that mutating Phe-54 reduced the amount of AC75 but did not affect the localization of AC75 in infected cells. A protein stability assay showed that the Phe-54 mutation affected AC75 stability. Taken together, Phe-54 was identified as an important residue of AC75, and *ac75* is a pivotal gene in budding virus production and occlusion body formation.

Keywords: Autographa californica multiple nucleopolyhedrovirus, AC75, Phe-54, budded virus, occlusion body

INTRODUCTION

Autographa californica multiple nucleopolyhedrovirus (AcMNPV), the type species of the genus *Alphabaculovirus* in the family of *Baculoviridae*, has a circular, double-stranded DNA genome of 134 kb that codes for 154 genes (Ayres et al., 1994). AcMNPV produces two viral forms: budded viruses (BVs) and occlusion-derived viruses (ODVs) (van Oers and Vlak, 2007; Rohrmann, 2019). BVs are responsible for spreading AcMNPV infection among susceptible insect cells and tissues (Blissard and Wenz, 1992), whereas ODVs initiate primary infection in the midgut epithelium of insects (Braunagel and Summers, 2007; Hodgson et al., 2007). AcMNPV genome replication and nucleocapsid assembly both occur in the nucleus. The synthesized nucleocapsids egress from the nucleus and bud from the plasma membrane to form mature BVs, whereas the retained

nucleocapsids are enveloped with intranuclear microvesicles to form ODVs that are further enclosed within polyhedrins to form occlusion bodies (OBs).

Baculovirus genome replication and transcription occur in the virogenic stroma (VS) (Young et al., 1993). The baculovirus genome is packaged into pre-assembled capsid sheaths to form mature nucleocapsids. Subsequently, a subset of the nucleocapsids egress from the nucleus to produce BVs. According to a previous report, AcMNPV nucleocapsid egress involves actin-based movement within nuclear envelope protrusions, which is coupled with localized nuclear envelope disruption and viral release into the cytoplasm (Ohkawa and Welch, 2018). Currently, several AcMNPV genes (*ac11*, *ac13*, *ac51*, *ac66*, *ac78, gp41*, *ac93*, *p48*, *exon0*, and *p49*) have been identified to be required for nuclear egress (Fang et al., 2007; Ke et al., 2008; McCarthy et al., 2008; Yuan et al., 2008, 2011; Tao et al., 2013, 2015; Biswas et al., 2017; Li et al., 2018; Qiu et al., 2019; Chen et al., 2020). Deletion of these genes does not affect viral genome replication and progeny nucleocapsid assembly but restrains or blocks the egress of progeny nucleocapsids from the nucleus to the cytoplasm. Recent research has shown that a group of AcMNPV proteins interact with components of the endosomal sorting complex required for transport-III complex and this complex is required for nuclear egress (Yue et al., 2018).

Nucleocapsids retained in nuclei are enveloped by intranuclear microvesicles to form ODVs and are further enclosed within the polyhedrins to form OBs (Rohrmann, 2019). The assembly and occlusion of ODVs, including intranuclear microvesicle formation, involves nucleocapsid bundles adhering to microvesicles, and subsequent wrapping of nucleocapsids by these microvesicles (Blissard and Theilmann, 2018). In previous reports, a number of AcMNPV genes (*ac11*, *ac76*, *ac93*, *odv-e25*, *p48*, and *p49*) have been shown to be involved in ODV envelopment (McCarthy et al., 2008; Yuan et al., 2008, 2011; Hu et al., 2010; Chen et al., 2012; Tao et al., 2015). Among the aforementioned genes, three genes (*ac11*, *ac93*, and *p48*) are required for both nuclear egress of the nucleocapsids and formation of intranuclear microvesicles (McCarthy et al., 2008; Yuan et al., 2008, 2011; Tao et al., 2015; Wang et al., 2019). The OBs form after the mature ODVs are embedded within polyhedrins. Then, a carbohydrate-composed outer layer and the polyhedral envelope protein (PEP) are supplemented to the surface of mature OBs (Whitt and Manning, 1988; Russell and Rohrmann, 1990), where a functional P10 protein associated with nuclear fibrillar structures is also required (Williams et al., 1989; Carpentier et al., 2008). In deletion mutants of *pep* or *p10*, OBs are irregular and fragile and no outer calyx layer is observed (Vlak et al., 1988; Li et al., 2015). Although several genes essential for OB morphogenesis have been reported, the detailed mechanism of OB morphogenesis is poorly understood.

ac75 is a highly conserved gene found in all sequenced baculovirus genomes, except Culex nigripalpus nucleopolyhedrovirus (CuniNPV), and *ac75* is predicted to encode a protein of 133 amino acids in length with a putative molecular mass of 15.5 kDa. AC75 localizes predominantly in the intranuclear ring zone in the late infection phase, while also exhibiting a nuclear rim distribution during the early phase (Guo et al., 2017; Shi et al., 2017). AC75 has been reported to be associated with the envelope and nucleocapsid fractions of BVs but only with the nucleocapsid fraction of ODVs (Shi et al., 2017). In contrast, another report revealed that AC75 is associated with the nucleocapsid of BVs and with both the envelope and nucleocapsid of ODVs (Guo et al., 2017). Nevertheless, two reports both confirmed that *ac75* was required for nucleocapsid egress and intranuclear microvesicle formation (Guo et al., 2017; Shi et al., 2017).

In this study, sequence alignment revealed that Phe-54 and Gln-81 of AC75 were highly conserved among alphabaculoviruses and betabaculoviurses. We constructed Phe-54 and Gln-81 point mutation bacmids (bAc^{ac75F54S}-*ph* and bAc^{ac75Q81A}-*ph*) and evaluated the effects of these mutations on virus proliferation. The data showed that the Gln-81 mutation did not affect viral propagation, whereas the Phe-54 mutation impaired BV production, polyhedrin assembly and ODV embedding. Thus, the results showed that Phe-54 is a key residue of AC75, and *ac75* is essential for OB morphogenesis.

MATERIALS AND METHODS

Cell Lines, Viruses, Insects, and Antibodies

Sf9 cells (Invitrogen, Carlsbad, CA, United States) were cultured at 27°C in Grace's insect medium (Invitrogen) supplemented with 10% (v/v) fetal bovine serum (Gibco, Grand Island, NY, United States) and 0.1% (v/v) antibiotic-antimycotic solution (Invitrogen). The recombinant bacmids were generated from bMON14272 (Invitrogen) and maintained in *Escherichia coli* (*E. coli*) strain DH10B (Invitrogen), which also contained helper plasmids for homologous recombination and transposition. *Spodoptera exigua* (*S. exigua*) larvae were reared on an artificial diet at 28°C (Shorey and Hale, 1965).

The polyclonal antiserum of anti-AC75 was prepared in rabbits according to previously published methods (Li et al., 2018). The anti-POLH polyclonal antiserum to detect the polyhedrin protein was preserved in the laboratory. Mouse monoclonal anti-actin antibody, horseradish peroxidase (HRP)-conjugated goat anti-mouse antibody and HRP-conjugated goat anti-rabbit antibody were purchased from Proteintech (Wuhan, China).

Construction of an *ac75* Knockout Bacmid

The AcMNPV *ac75* gene was deleted using the λ red recombination system in *E. coli* BW25113 cells (containing bMON14727 and pKD46) as described previously (Datsenko and Wanner, 2000; Li et al., 2015). The *ac75*-null bacmid was constructed by replacing a 112-bp fragment of the *ac75* ORF with a chloramphenicol resistance gene (*CmR*) cassette and retaining 115 nt of the 5′-end and 175 nt of the 3′-end of the *ac75* ORF to avoid affecting transcription of neighboring genes *ac76* and *ac74*. The *ac75* knockout bacmid was verified by PCR (primer sequences are shown in **Table 1**) and termed bAc^{ac75KO}.

TABLE 1 | Primers used in this study.

Primer name	Primer sequence (5′-3′)[a]
ac75-US-F (*Sac*I)	CGAGCTCCGCAACGAATAGAGTAAGGG
ac75-US-R (*Bam*HI)	CGGGATCCGATAGACTTGTTCGCACAGC
CmR-F (*Bam*HI)	CGGGATCCTGTAGGCTGGAGCTGC
CmR -R (*Hin*dIII)	CCCAAGCTTCATATGAATATCCTCCTTAGTTCC
ac75-DS-F (*Hin*dIII)	CCCAAGCTTACTCAGTAGGCGACAGGTTG
ac75-DS-R (*Xho*I)	CCGCTCGAGCTTTGGCGTGGTCAATG
ph-F (*Eco*RI)	CGGAATTCACCATCTCGCAAATAAATAAG
ph-R (*Sac*I)	CGAGCTCTGTATCGTGTTTTAATACGCC
egfp-F (*Sma*I)	CCCCGGGATGGTGAGCAAGGGCGAGGAGC
egfp-R (*Xho*I)	CCGCTCGAGTCACTTGTACAGCTCGTCCATGCCGAG
Dual-*ac75*-F1	CCGGAGTAGCCATATTTAGCCTAGTGTATGAC
Dual-*ac75*-R1	CAGAATTCTTAATACGCTGGCAGTTGGTATG
Dual-*ac75*-R2	TTACTTATCGTCGTCATCCTTGTAATCATACGCTGGCAG TTGGTATG
pFast-*ac75*-F1	CGTATTAAGAATTCTGCAGATATCCAGCAC
pFast-*ac75*-F2	TGACGACGATAAGTAAGAATTCTGCAGATATCCAGCAC
pFast-*ac75*-R1	AATATGGCTACTCCGGAATATTAATAGATCATGGAG
$ac75^{F54S}$-F	AACTCAAACGAGTAGTTAACATGAGTTTAAACAATG
$ac75^{F54S}$-R	TGTTTAAACTCATGTTAACTACTCGTTTGAGTTTAAGC
$ac75^{Q81A}$-F	AGTAGGCGAGCGGTTGATTTTTTAATACATG
$ac75^{Q81A}$-R	AAATCAACCGCTCGCCTACTGAGTTTATTAG
ac75-F	ATGTCCAATTTAATGAAAAACTTTTTCACC
ac75-R	TTAATACGCTGGCAGTTGGTATGC
qie1-F	TGTGATAAACAACCCAACGAC
qie1-R	GTTAACGAGTTGACGCTTG
qpe38-F	AATGGAACAGCAGCGAATGA
qpe38-R	GTCGCACGTAGTCGGAATC
qgp64-F	ACGACCTGATAGTCTCCGTG
qgp64-R	TGTAGCAATTACTGGTGTGTGC
qvp39-F	TTGCGCAACGACTTTATACC
qvp39-R	TAGACGGCTATTCCTCCACC
qpolh-F	TTAGGTGCCGTTATCAAGA
qpolh-R	GCCACTAGGTAGTTGTCT
q18S-F	TACCGATTGAATGATTTAGTGAGG
q18S-R	TACGGAAACCTTGTTACGACTTT
pIB-F1	GTCCAGTGTGGTGGAATTCTG
pIB-F2	CGGCGGCAGCGGCGGCGGCAGCCCCGGGATGGTG AGCAAGGGCGAGGAGC
pIB-R	TAGTGGATCCGAGCTCGGTAC
pIB-*egfp*-F	GAGCTCGGATCCACTAATGGTGAGCAAGG GCGAGGAGC
pIB-*egfp*-R	TTCCACCACACTGGACCTACTTGTACAGCTCGTC CATGCCGAG
pIB-*ac75*-F	GAGCTCGGATCCACTAATGTCCAATTTAATGAAAAA CTTTTTCACC
pIB-*ac75*-R	CGCCGCTGCCGCCGCCATACGCTGGCAGTTGGTATGC

[a] *Restriction sites and homologous sequences were underlined.*

Construction of *ac75* Recombinant Bacmids

The AcMNPV *ac75* gene knockout and repair bacmids were generated as described previously (Chen et al., 2020). Briefly, the *polh* and *egfp* genes were, respectively, cloned downstream of the *polh* and *p10* gene promoters in pFastBacDual to generate donor plasmid pFBD-*ph-egfp*. The fragment containing the *ac75* promoter and ORF was inserted into pFBD-*ph-egfp* to give the recombinant plasmid pFBD-*ph-ac75-egfp*. Subsequently, the donor plasmids pFBD-*ph-egfp* and pFBD-*ph-ac75-egfp* were transformed into *E. coli* DH10B competent cells (containing the bacmid bAc^{ac75KO} and a helper plasmid). Recombinant bacmids were selected by gentamicin and kanamycin resistance with blue-white screening and further identified by PCR (primer sequences are shown in **Table 1**).

The *ac75* point mutation donor plasmids were constructed using FastCloning (Li et al., 2011) and the corresponding recombinant bacmids were generated via the bac-to-bac system. For example, to generate the AC75-F54S mutant donor plasmid, the fragment was amplified by using the primer pairs $ac75^{F54S}$-F/R (sequences are shown in **Table 1**) and the template pFBD-*ph-ac75-egfp*. Then, 1 μL of the enzyme *Dpn*I (Takara, Toyoko, Japan) was added to the PCR product (9 μL) and digested at 37°C for 1 h. The mixture was transformed into *E. coli* DH5α competent cells. The recombinant plasmid selected by gentamicin and ampicillin resistance was further validated by PCR and DNA sequencing. The donor plasmid was termed pFBD-*ph-ac75*F54S-*egfp*. Subsequently, the mutation plasmid pFBD-*ph-ac75*F54S-*egfp* was further transformed into DH10B competent cells (containing the bAc^{ac75KO} bacmid and a helper plasmid) to generate mutation bacmid $bAc^{ac75F54S}$-*ph* via the bac-to-bac system. The point mutation recombinant bacmid of $bAc^{ac75Q81A}$-*ph* was also constructed by the same methods.

Transmission Electron Microscopy and Immunoelectron Microscopy Analyses of Virus Infected Cells

Transmission electron microscopy (TEM) analysis was performed according to previous results (Qin et al., 2019). Sf9 cells infected with vAc-*ph*, $vAc^{ac75F54S}$-*ph*, or $vAc^{ac75REP}$-*ph* at an MOI of 5 were fixed with 2.5% (v/v) glutaraldehyde for 2 h and harvested at 24, 36, and 48 h post infection (p.i.). Ultrathin sections were visualized by a FEI Tecnai G^2 20 TWIN TEM. For immunoelectron microscopy analysis, Sf9 cells were infected with vAc-*ph*, $vAc^{ac75F54S}$-*ph* or $vAc^{ac75REP}$-*ph* at an MOI of 5. At 48 h p.i., the cells were fixed with 1% paraformaldehyde-0.5% glutaraldehyde for 10 min at 4°C, refixed with 2% paraformaldehyde-2.5% glutaraldehyde for 1 h at 4°C and then dehydrated and embedded according to previous methods (Xu et al., 2019). Ultrathin sections were immunostained with anti-AC75 pAb (1:50) as the primary antibody. Goat anti-rabbit IgG coated with gold particles (10 nm; Sigma, Darmstadt, Germany) was used as the secondary antibody (1:50). Ultrathin sections were also visualized by using the FEI Tecnai G^2 20 TWIN TEM.

Scanning Electron Microscopy (SEM), TEM, and Negative Staining Analyses of OBs

OBs were amplified and isolated from the infected larvae according to the method described by Gross et al. (Gross et al., 1994). For SEM, OBs (10^8 OBs/mL) were loaded onto the silver paper and dried at 37°C overnight. The samples were sputter

coated with gold and examined by SEM (Hitachi SU8010). For TEM, OBs were fixed with 2.5% (v/v) glutaraldehyde for 2 h and prepared as described previously (Ji et al., 2015). Ultrathin section images were obtained by TEM (FEI Tecnai G^2 20 TWIN). For negative staining analysis, 10 μL of an OB suspension (10^8 OBs/mL) was loaded onto a copper grid for 10 min. Filter paper was used to remove the remaining solution from the grid. Then, 10 μL dissolution buffer was added to dissolve the OBs for 1 min. After removing the dissolution buffer, the grid was stained with 2% (w/v) phosphotungstic acid (pH 5.7) for 1 min. The grid was also examined by TEM. The number of OBs with ODVs in a field of view were counted using ImageJ software[1], and the numbers were analyzed using the Kruskal-Wallis test followed by Dunn's multiple comparison test.

Immunofluorescence Microscopy

To investigate whether AC75-F54S affected the localization of AC75, immunofluorescence assays were performed as described previously with minor modifications (Hepp et al., 2018). Briefly, Sf9 cells were seeded (4×10^5 cells/dish) on a glass dish and allowed to attach for 2 h. The cells were infected with vAc-*ph*, vAc^{ac75F54S}-*ph* or vAc^{ac75REP}-*ph* at an MOI of five and fixed with 4% paraformaldehyde for 10 min at 15, 24, and 48 h p.i. After treatment with 0.2% (v/v) Triton X-100 for 10 min and blocked with PBS containing 5% (w/v) bovine serum albumin and 0.1% (v/v) Tween-20 for 30 min, the cells were incubated with rabbit anti-AC75 polyclonal antiserum (1:500) for 1 h followed by Alexa Fluor 594 goat anti-rabbit IgG (1:1000, Invitrogen) for 1 h in the dark. Subsequently, the cell nuclei were stained with Hoechst 33258 (Beyotime, Shanghai, China). All samples were observed with a 60× oil-immersion objective and a PerkinElmer UltraView VOX system.

Quantitative Analysis of Viral Gene Transcription

Sf9 cells (1.0×10^6 cells/plate) were transfected in triplicate with bAc^{ac75F54S}-*ph* or bAc^{ac75REP}-*ph* bacmid DNA and collected at 24 h post-transfection (p.t.). Total cellular RNA was isolated by using RNAiso Plus (Takara). The cDNA was then synthesized using an iScript cDNA synthesis kit (Takara). qPCR was performed with five pairs of specific viral gene primers (sequences are shown in **Table 1**) by a CFX96 real-time system (Bio-Rad) under the following conditions: denaturation at 95°C for 3 min followed by 40 cycles of 95°C for 10 s, 55°C for 30 s and 72°C for 20 s. Melting curve analysis was performed at the end of each PCR assay to test specificity (control). Host 18S rRNA was selected and used as the endogenous reference.

Western Blot Analysis

Analysis of AC75 expression in cells infected with vAc^{ac75F54S}-*ph* or vAc^{ac75REP}-*ph* over the course of infection was performed by western blot. Sf9 cells were infected in triplicate with vAc^{ac75F54S}-*ph* or vAc^{ac75REP}-*ph* at an MOI of 5 and harvested at 18 and 24 h p.i. The cells were lysed on ice for 10 min in lysis buffer (Beyotime) and centrifuged in a microcentrifuge at 16,000 × *g* for 2 min. The lysates were subjected to western blot as described previously with some modifications (Xu et al., 2019). Anti-AC75 and anti-actin were used as primary antibodies. The signal was detected using a BeyoECL Plus Kit (Beyotime).

Protein Stability Analysis

Protein stability was determined by using a previously described assay (Byers et al., 2016). Briefly, Sf9 cells were infected with vAc^{ac75F54S}-*ph* or vAc^{ac75REP}-*ph* at an MOI of five and were incubated with medium containing 400 μg/mL cycloheximide (CHX; Sigma) at 24 h p.i. The cells were harvested and lysed at the designated times, and the lysates were analyzed by western blot.

RESULTS

ac75 Is a Late Viral Gene

Transcriptomic analysis showed that three late transcription start sites are located upstream of the *ac75* translation initiation codon (Chen et al., 2013). Reverse transcription (RT)-PCR was performed in AcMNPV infected cells to confirm the temporal transcription patterns of *ac75*. The *ac75* transcripts were detected from 12 h p.i. and persisted up to 72 h p.i. (**Figure 1A**). Furthermore, western blot analysis indicated that aphidicolin inhibited AC75 expression (**Figure 1B**). These results confirmed that *ac75* was a late viral gene, implying its possible role during the late life cycle of AcMNPV.

Phe-54 Is a Key Residue of AC75

According to previous reports, *ac75* is an essential gene in the life cycle of the virus (Guo et al., 2017; Shi et al., 2017). Cells transfected with the *ac75*-null bacmid could not produce progeny BVs, which precluded analysis of the roles that AC75 plays throughout the life cycle of the virus. An InterProScan (Jones et al., 2014) and NCBI Conserved Domain Search (Marchler-Bauer et al., 2017) indicated that only a functionally unknown DUF1160 constitutes AC75, with no other annotated domains. To further investigate critical residues and functions of AC75, the sequences of AC75 homologs from 59 alphabaculoviruses and 14 betabaculoviruses were aligned by Clustal X-2.0 (Larkin et al., 2007). Residues Phe-54 and Gln-81 were found to be completely conserved among the homologs in the selected baculoviruses (**Supplementary Figure 1**). Based on this observation, we hypothesized that Phe-54 and Gln-81 from AC75 play key roles. Consequently, Phe-54 and Gln-81 point mutation bacmids (bAc^{ac75F54S}-*ph* and bAc^{ac75Q81A}-*ph*) were constructed (**Figure 2**). Sf9 cells were transfected with bAc^{ac75F54S}-*ph*, bAc^{ac75Q81A}-*ph*, bAc^{ac75REP}-*ph* or bAc-*ph* and monitored by fluorescent microscopy. There were no significant differences in the number of fluorescent cells among bacmids at 24 h p.t. (**Figure 3A**, upper panel), indicating relatively equal frequencies of transfection of these bacmids. By 72 h p.t., most of the bAc^{ac75Q81A}-*ph*-, bAc^{ac75REP}-*ph*- or bAc-*ph*-transfected cells exhibited fluorescence, whereas only a small proportion of bAc^{ac75F54S}-*ph*-transfected cells showed fluorescence (**Figure 3A**, middle panel). In addition, light

[1]https://imagej.nih.gov/ij/

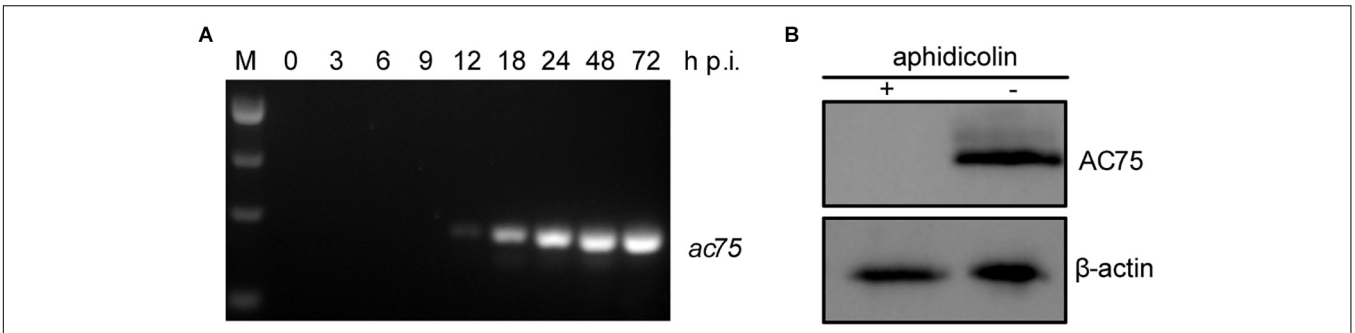


FIGURE 1 | Analyses of the transcription and expression profile of *ac75* in infected cells. **(A)** Time course analysis of *ac75* transcription. Sf9 cells infected with AcMNPV were collected at the indicated time points. Total RNA was extracted and total cDNA was obtained by reverse transcription. *ac75* was amplified with specific primer pairs by PCR. **(B)** Western blot analysis of the expression of AC75 following aphidicolin treatment. AcMNPV infected cells were treated with 5 μg/mL aphidicolin (+) or DMSO (–) at 0 h p.i. The cells were lysed and subjected to western blot analysis with anti-AC75 and anti-actin as primary antibodies at 24 h p.i.

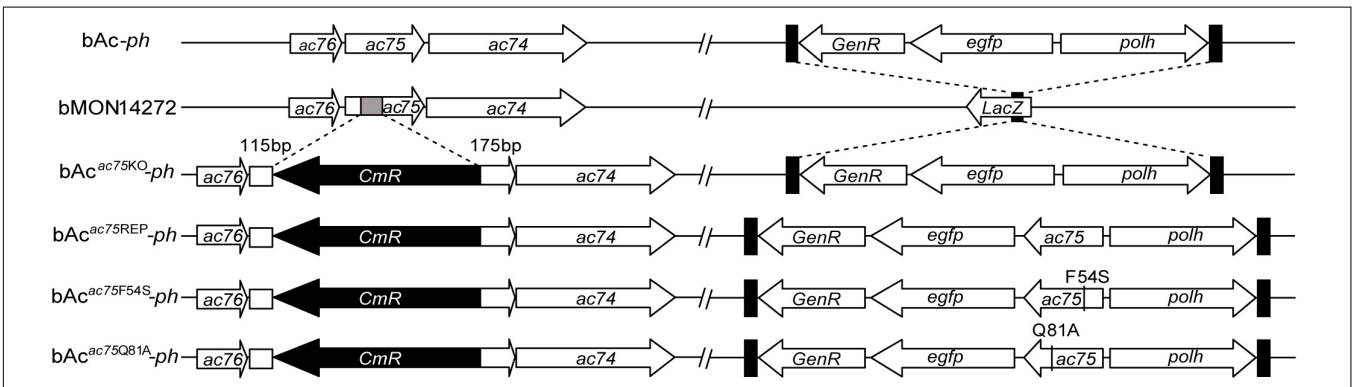


FIGURE 2 | Schematic diagram of AcMNPV, *ac75* knockout, repaired and point mutation bacmids. The 112-bp fragment of *ac75* ORF in bMON14272 was replaced with a *CmR* cassette via homologous recombination to generate bAc^{ac75KO}. bAc-*ph* and bAc^{ac75KO}-*ph* were generated by inserting the *polh* and *egfp* genes into the *polh* locus of bMON14272 and bAc^{ac75KO}, respectively. The *ac75*, *ac75*F54S or *ac75*Q81A controlled by the *ac75* native promoter with the *polh* and *egfp* genes were inserted into the *polh* locus of bAc^{ac75KO} to generate bAc^{ac75REP}-*ph*, bAc^{ac75F54S}-*ph* and bAc^{ac75Q81A}-*ph*, respectively.

microscopy images revealed that there were fewer intracellular OBs in bAc^{ac75F54S}-*ph*-transfected cells when compared with cells transfected with the other bacmids (**Figure 3A**, lower panel), suggesting that the AC75-F54S mutation may also affect OB production. Furthermore, virus growth curve analysis showed that bAc^{ac75Q81A}-*ph*, bAc^{ac75REP}-*ph* and bAc-*ph* had comparable growth kinetics, whereas bAc^{ac75F54S}-*ph* produced fewer progeny BVs from 24 to 120 h p.t. (**Figure 3B**). Taken together, these results revealed that Phe-54 is a key residue of AC75, and the AC75-F54S mutation may affect progeny BV production and OB formation.

To further confirm the results obtained from bacmid transfection, Sf9 cells were infected with vAc-*ph*, vAc^{ac75F54S}-*ph*, vAc^{ac75Q81A}-*ph* or vAc^{ac75REP}-*ph* at an MOI of 0.5, and virus growth curve analysis was performed. At 96 h p.i., the number of fluorescent cells infected with mutant virus vAc^{ac75F54S}-*ph* was less than those infected with vAc-*ph*, vAc^{ac75Q81A}-*ph* or vAc^{ac75REP}-*ph* (**Figure 3C**, upper panel), and the number of OBs produced by vAc^{ac75F54S}-*ph*-infected cells was also less when compared with those cells infected with vAc-*ph*, vAc^{ac75Q81A}-*ph* or vAc^{ac75REP}-*ph* (**Figure 3C**, lower panel). The BV production of the vAc^{ac75F54S}-*ph* virus was reduced by 10-fold when compared with those of vAc-*ph*, vAc^{ac75Q81A}-*ph* or vAc^{ac75REP}-*ph* from 24 to 96 h p.i. (**Figure 3D**). The number of OBs in cells infected with vAc-*ph*, vAc^{ac75F54S}-*ph*, vAc^{ac75Q81A}-*ph*, or vAc^{ac75REP}-*ph* at an MOI of 10 were counted. The OBs of vAc^{ac75F54S}-*ph*-infected cells from each dish were reduced by approximately eight-fold when compared with those cells infected with vAc-*ph*, vAc^{ac75Q81A}-*ph* or vAc^{ac75REP}-*ph* at 96 h p.i. ($P < 0.001$) (**Figure 3E**). These observations validated that the AC75-F54S mutation affected BV production and OB formation.

The AC75-F54S Mutation Affected Polyhedrin Assembly and ODV Embedding

Transmission electron microscopy was performed to further determine the effects of the AC75-F54S mutation on intranuclear structures. At 24 and 36 h p.i., vAc^{ac75F54S}-*ph*-infected cells exhibited typical baculovirus infection symptoms, including a typical VS and abundant rod-shaped nucleocapsids (**Figure 4B**), and ODVs with nucleocapsids (**Figure 4E**). As expected, vAc-*ph*- and vAc^{ac75REP}-*ph*-infected cells also showed a typical VS and normal rod-shaped nucleocapsids (**Figures 4A,C**), and

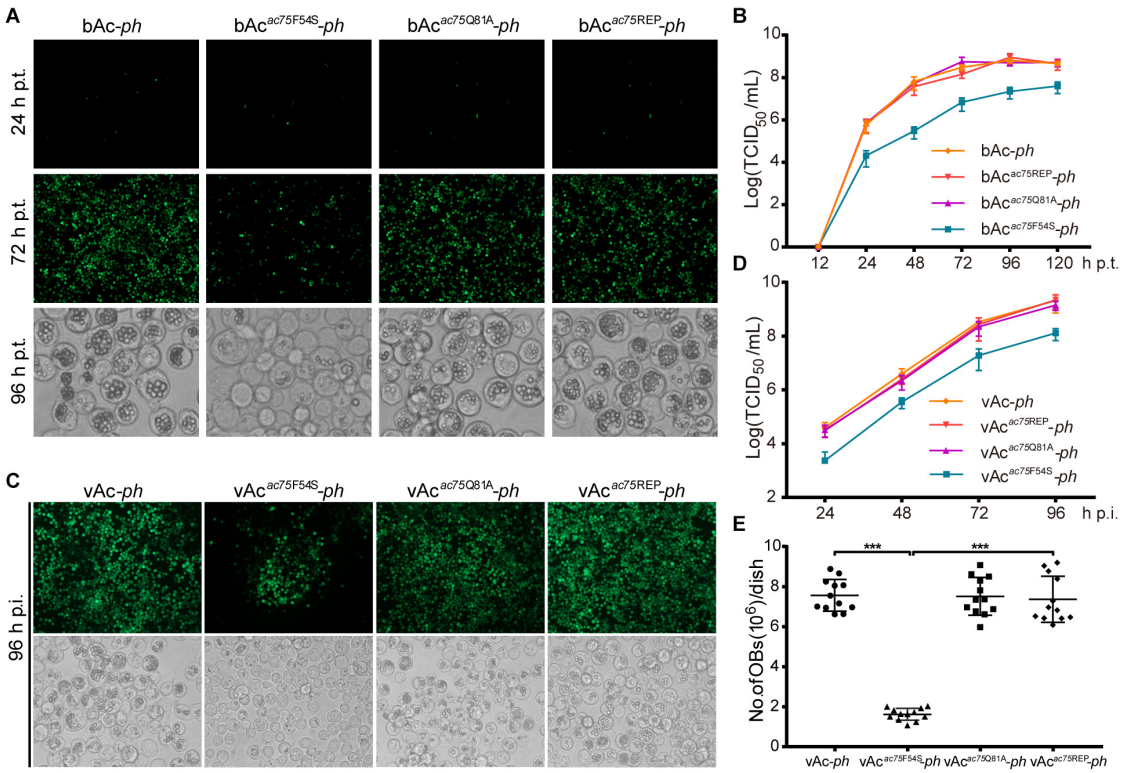


FIGURE 3 | Analyses of viral replication and OB formation in the transfected/infected cells. **(A)** Fluorescence and light microscopy analyses of bacmid-transfected cells. Sf9 cells were transfected with bAc-*ph*, bAc^{ac75F54S}-*ph*, bAc^{ac75Q81A}-*ph* or bAc^{ac75REP}-*ph* and imaged at 24, 72, and 96 h p.t. **(B)** Virus growth curves generated from bacmid-transfected cells. Sf9 cells were transfected with bAc-*ph*, bAc^{ac75F54S}-*ph*, bAc^{ac75Q81A}-*ph*, or bAc^{ac75REP}-*ph*. The supernatants were harvested at the indicated time points, and virus titers were determined by the endpoint dilution assay. Each data point represents average titers from three separate transfections. Error bars represent standard deviations (SD). **(C)** Fluorescence and light microscopy analyses of virus-infected cells. Sf9 cells were infected with vAc-*ph*, vAc^{ac75F54S}-*ph*, vAc^{ac75Q81A}-*ph* or vAc^{ac75REP}-*ph* (MOI = 0.5) at 96 h p.i. **(D)** Virus growth curves generated from virus-infected cells. Sf9 were infected with vAc-*ph*, vAc^{ac75F54S}-*ph*, vAc^{ac75Q81A}-*ph* or vAc^{ac75REP}-*ph* at an MOI of 0.5. The supernatants were harvested at the designated times and determined by the endpoint dilution assay. Each data point represents average titers from three separate infections. Error bars represent SD. **(E)** Amount of OB production in each dish. Sf9 cells infected vAc-*ph*, vAc^{ac75F54S}-*ph*, vAc^{ac75Q81A}-*ph* or vAc^{ac75REP}-*ph* (MOI = 10) were harvested at 96 h p.i. Total OBs of each dish were measured using a hemocytometer (*** indicates $P < 0.001$).

numerous ODVs (**Figures 4D,F**) at 24 and 36 h p.i. However, the OBs of vAc^{ac75F54S}-*ph*-infected cells contained fewer envelope virions (**Figure 4H**) when compared with those of vAc-*ph* and vAc^{ac75REP}-*ph*-infected cells (**Figures 4G,I**) at 48 h p.i., indicating that the AC75-F54S mutation impaired OB occlusion ODVs.

To further investigate the effects of the AC75-F54S mutation on OB morphogenesis, we performed SEM and TEM on OBs purified from vAc-*ph*-, vAc^{ac75F54S}-*ph*- or vAc^{ac75REP}-*ph*-infected *S. exigua* cadavers. SEM analysis showed that the OBs from vAc^{ac75F54S}-*ph*-infected larvae had ragged surfaces and irregular shapes, whereas the OBs of vAc-*ph* or vAc^{ac75REP}-*ph*-infected larvae had smooth surfaces and sharp edges (**Figure 5**, upper panel). TEM images showed that only a few OBs of vAc^{ac75F54S}-*ph* contained ODVs, whereas the majority of OBs of vAc-*ph* and vAc^{ac75REP}-*ph* occluded normal ODVs with multi-nucleocapsids (**Figure 5**, middle panel). Furthermore, twenty visual fields were randomly selected and analyzed, the proportion of vAc^{ac75F54S}-*ph* OBs with ODVs were only 7.8%. In addition, OBs were dissolved, negative stained and examined by TEM. Many OBs of vAc^{ac75F54S}-*ph* were empty (**Figure 5**, lower panel), which is consistent with the observations of thin sections of OBs, whereas OBs of vAc-*ph* and vAc^{ac75REP}-*ph* contained normal single-nucleocapsids and multi-nucleocapsids (**Figure 5**, lower panel). According to these results, the AC75-F54S mutation caused aberrant polyhedrin assembly and ODV embedding during OB formation.

The AC75-F54S Mutation Did Not Affect Polyhedrin Expression and Localization

According to previous reports, *ac75* is essential in the life cycle of the virus but deleting *ac75* did not affect viral genome replication (Guo et al., 2017; Shi et al., 2017). To investigate whether the AC75-F54S mutation affected transcription of viral genes, five viral genes, including two early genes (*ie1* and *pe38*), one early-late gene (*gp64*) and two late genes (*vp39* and *polh*), were selected and analyzed. The transcript levels of the selected genes were determined by RT-qPCR with the corresponding primers (sequences are shown in **Table 1**). As shown in **Figure 6A**, no significant differences in transcript levels of all selected

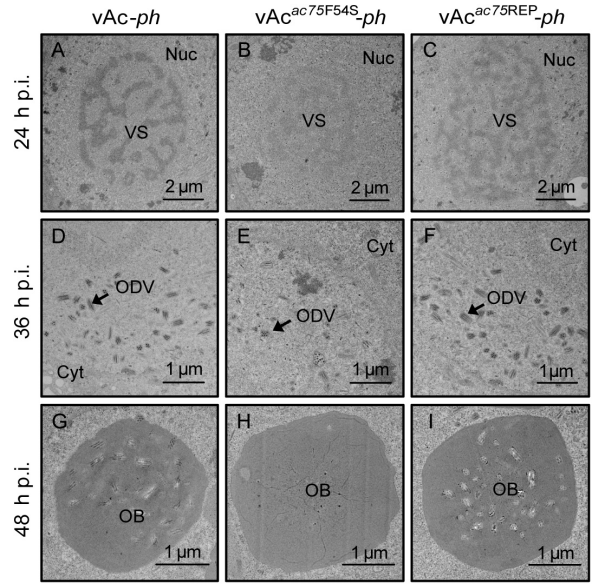


FIGURE 4 | TEM analysis of cells infected with vAc-*ph*, vAc^{ac75F54S}-*ph*, or vAc^{ac75REP}-*ph*. Sf9 cells were infected with vAc-*ph*, vAc^{ac75F54S}-*ph* or vAc^{ac75REP}-*ph*. The cells were collected at 24, 36, and 48 h p.i., and prepared for TEM. **(A–C)** The enlarged nucleus (Nuc) and virogenic stroma (VS) in vAc-*ph*, vAc^{ac75F54S}-*ph* or vAc^{ac75REP}-*ph*-infected cells at 24 h p.i. **(D–F)** The normal ODVs in the nucleus ring zone of cells infected with vAc-*ph*, vAc^{ac75F54S}-*ph* or vAc^{ac75REP}-*ph* at 36 h p.i. Cyt, cytoplasm. **(H)** The OB with a few ODVs in vAc^{ac75F54S}-*ph*-cells at 48 h p.i. **(G,I)** Normal OBs in vAc-*ph* or vAc^{ac75REP}-*ph*-infected cells at 48 h p.i.

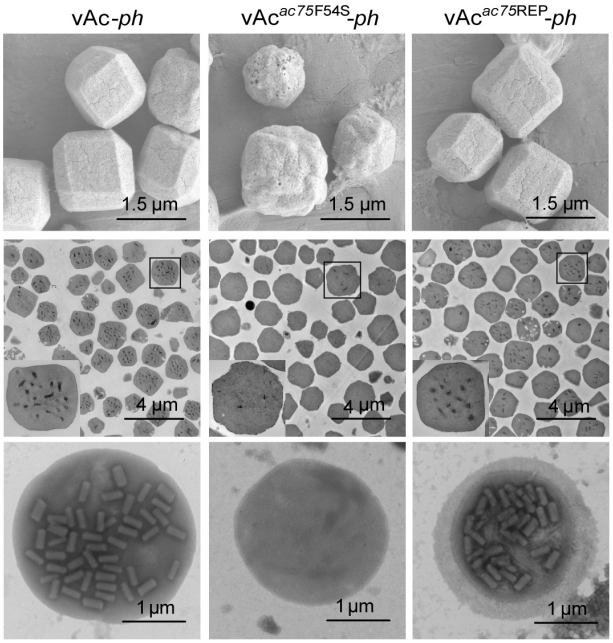


FIGURE 5 | SEM, TEM and negative staining analyses of OBs. The OBs were purified from *S. exigua* larvae infected with vAc-*ph*, vAc^{ac75F54S}-*ph* or vAc^{ac75REP}-*ph*. (Upper panel) SEM analyses of OBs. The images show the external morphology of OBs. (Middle panel) TEM analyses of OBs. The ultrathin sections enabled visualization of the internal structure of OBs. Boxed regions of the OB are magnified and shown in the inset. (Lower panel) Negative staining analyses of OBs.

genes were observed between bAc^{ac75F54S}-*ph*- and bAc^{ac75REP}-*ph*-transfected cells ($P > 0.05$). The results showed that the AC75-F54S mutation did not affect early or late viral gene transcription.

Western blot analysis was performed to determine whether the AC75-F54S mutation affected expression of the polyhedrin protein because the AC75-F54S mutation was found to impair OB formation. At 24 h p.i., there were no difference among the polyhedrin expression levels of vAc^{ac75F54S}-*ph*- and vAc^{ac75REP}-*ph*-infected cells (**Figures 6B,C**), indicating that the AC75-F54S mutation also did not affect polyhedrin expression. Furthermore, an immunofluorescence assay was performed to investigate the effect of the AC75-F54S mutation on polyhedrin localization. As shown in **Figure 6D**, polyhedrins were distributed predominantly in the nucleus, exhibiting a ring pattern in the ring zone, with a weak distribution in the cytoplasm of vAc-*ph*, vAc^{ac75F54S}-*ph*- and vAc^{ac75REP}-*ph*-infected cells at 24 h p.i. The results indicated that the AC75-F54S mutation did not affect the localization of the polyhedrin. Taken together, these data suggested that the AC75-F54S mutation might impair OB formation by affecting polyhedrin assembly.

The AC75-F54S Mutation Caused a Decreased in the Amount of AC75

To further investigate the nature of the defect caused by the AC75-F54S mutation, the relative protein expression levels of AC75 in vAc^{ac75F54S}-*ph*- or vAc^{ac75REP}-*ph*-infected cells were compared over the course of infection by western blot analysis. As shown in **Figures 7A**, AC75 was detected in both vAc^{ac75F54S}-*ph*- and vAc^{ac75REP}-*ph*-infected cells at 18 and 24 h p.i.; however, the protein band representing AC75 in vAc^{ac75F54S}-*ph*-infected cells was much weaker in intensity when compared with that of vAc^{ac75REP}-*ph*-infected cells. The protein expression levels were further examined by standardization against actin expression and densitometry analysis. The relative AC75 protein expression in vAc^{ac75REP}-*ph*-infected cells was 0.58 ± 0.19 and 1.00 ± 0.16 at 18 and 24 h p.i., respectively (**Figure 7B**). In contrast, the relative AC75 protein expression in vAc^{ac75F54S}-*ph*-cells was 0.15 ± 0.07 and 0.41 ± 0.07 at 18 and 24 h p.i., respectively (**Figure 7B**). Immunoelectron microscopy analysis was performed with thin sections generated from cells infected with vAc-*ph*, vAc^{ac75F54S}-*ph* or vAc^{ac75REP}-*ph*, and fixed at 48 h p.i. In agreement with the above results, the number of colloidal-gold-labeled AC75 in the vAc^{ac75F54S}-*ph*-infected cells was remarkably lower when compared with that of vAc-*ph*- or vAc^{ac75REP}-*ph*-infected cells (**Supplementary Figure 2**).

Furthermore, immunofluorescence assays were performed to investigate the effect of the AC75-F54S mutation on the localization and abundance of AC75. At 15 h p.i., the AC75 signal was distributed uniformly throughout cells infected with vAc^{ac75F54S}-*ph* or vAc^{ac75REP}-*ph* (**Figure 7C**). By 24 and 48 h p.i., AC75 was localized predominantly within the nucleus, particularly around the nuclear rim and became condensed to

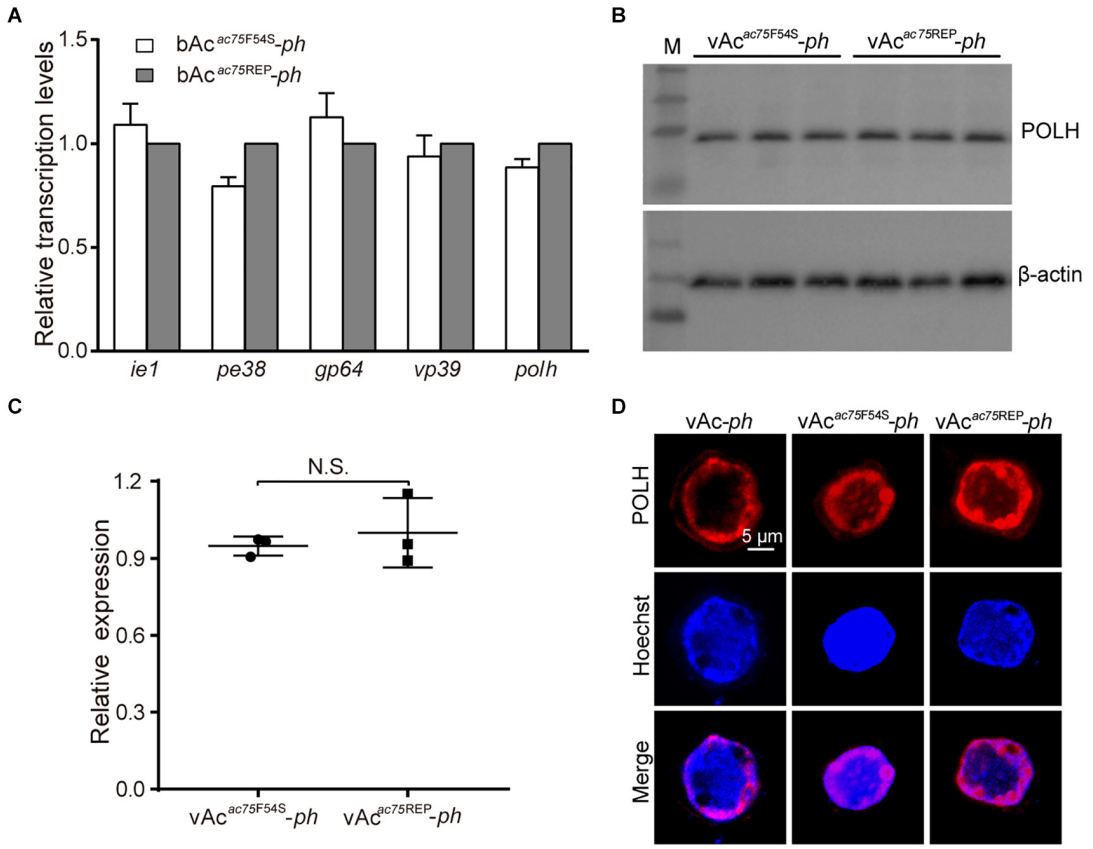


FIGURE 6 | Transcription, expression and localization analyses of *polyhedrin*. **(A)** RT-qPCR analysis. Total RNA was extracted from Sf9 cells transfected with either bAc^{ac75F54S}-*ph* or bAc^{ac75REP}-*ph* at 24 h p.t. Transcription of selected viral genes was measured by RT-qPCR. The transcript level of each gene was normalized by 18S rRNA of the same sample. The values represent the averages from three independent assays and error bars represent SD. **(B)** Western blot analysis. Sf9 cells infected with vAc^{ac75F54S}-*ph* or vAc^{ac75REP}-*ph* at an MOI of five were collected and lysed at 24 h p.i. The lysates were subjected to western blot analysis using anti-POLH and anti-actin (control) as primary antibodies. **(C)** Relative quantification of the expression levels of POLH by densitometry analysis of western blot. Each dot represents the expression level in one independent experiment relative to the mean for POLH in vAc-*ph*-infected cells at 24 h p.i. The values represent the averages from three independent assays and error bars represent SD. N.S. indicates no significance, $P > 0.05$. **(D)** Immunofluorescence analysis. Sf9 cells infected with vAc-*ph*, vAc^{ac75F54S}-*ph* or vAc^{ac75REP}-*ph* at an MOI of 5 were fixed at 24 h p.i. and stained. The POLH are in red and nuclei are stained blue.

the intranuclear ring zone in vAc^{ac75F54S}-*ph*- and vAc^{ac75REP}-*ph*-infected cells (**Figure 7C**). Thus, the localization of AC75 was similar in vAc^{ac75F54S}-*ph*- and vAc^{ac75REP}-*ph*-infected cells. In contrast, the AC75 signal in vAc^{ac75F54S}-*ph*-infected cells was much weaker when compared with those in vAc^{ac75REP}-*ph*- and vAc-*ph*-infected cells at all-time points (**Figure 7C**). The localization and abundance of AC75 in vAc-*ph*-infected cells were comparable with those of vAc^{ac75REP}-*ph* (**Supplementary Figure 3**). Taken together, these data demonstrated that the AC75-F54S mutation did not affect localization of AC75 in infected cells but caused a reduction in the amount of AC75.

The AC75-F54S Mutation Affected the Stability of AC75

The decrease in the amount of AC75 is probably because translation or stability of AC75 is affected by the AC75-F54S mutation. We initially examined and compared the expression of AC75 in cells transfected with pIB-*ac75^{F54S}egfp* or pIB-*ac75egfp* plasmids to determine whether the translation or the stability of AC75 was affected. The AC75 signal of cells transfected with pIB-*ac75^{F54S}egfp* was much weaker than that of pIB-*ac75egfp* with the same exposure time at 48 h p.t. (**Figure 8A**). Subsequently, we collected the cells and analyzed the mean fluorescence intensity of cells with EGFP by flow cytometry. As shown in **Figure 8B**, the mean fluorescence intensity of cells transfected with pIB-*ac75^{F54S}egfp* was significantly lower than that of pIB-*ac75egfp* ($P < 0.001$). These results indicated that the AC75-F54S mutation might impair AC75 stability.

To confirm the above results, a cycloheximide (CHX) treated assay was used to evaluate the relative stability of AC75. Briefly, Sf9 cells infected with vAc^{ac75F54S}-*ph* or vAc^{ac75REP}-*ph* at an MOI of five were treated with 400 μg/mL CHX to block protein synthesis at 24 h p.i. The cells were collected and lysed at 0, 3, 6 and 12 h after CHX treatment, and the lysates were analyzed by western blot. The AC75 protein level in vAc^{ac75F54S}-*ph*-infected cells was reduced significantly from 0 to 12 h (**Figure 8C**). In contrast, there were no obvious changes in AC75 protein

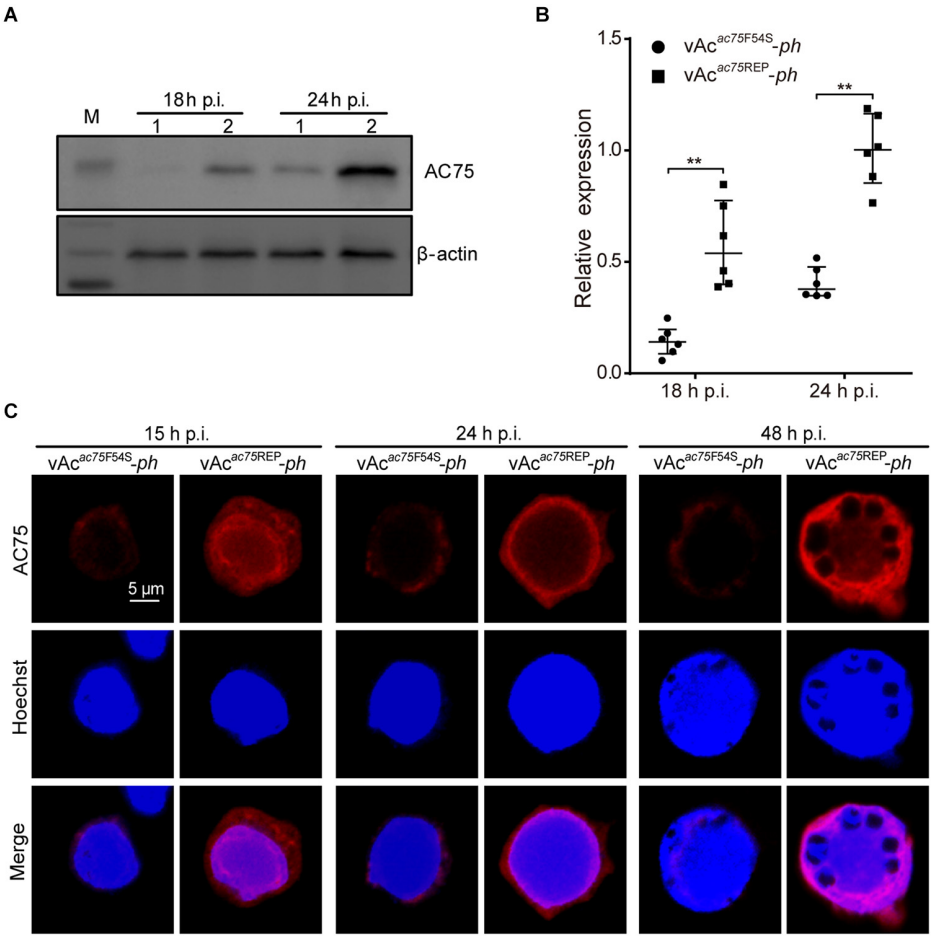


FIGURE 7 | Expression and localization of AC75 in vAc^{ac75F54S}-*ph*- or vAc^{ac75REP}-*ph*-infected cells. **(A)** Western blot analysis. Sf9 cells infected with vAc^{ac75F54S}-*ph* or vAc^{ac75REP}-*ph* at an MOI of 5 were harvested and lysed at 18 and 24 h p.i. The lysates were subjected to western blot using anti-AC75 and anti-actin (control) as primary antibodies. Lane 1, vAc^{ac75F54S}-*ph*; Lane 2, vAc^{ac75REP}-*ph*. **(B)** Relative quantification of the expression levels of AC75 by densitometry analysis of western blot. Each dot represents the expression level in one independent experiment relative to the mean for AC75 in vAc^{ac75REP}-*ph*-infected cells at 24 h p.i. The values represent the averages from six independent assays and error bars represent SD. ** indicates $P < 0.01$. **(C)** Immunofluorescence analysis of localization and expression of AC75. Sf9 cells infected with vAc^{ac75F54S}-*ph* or vAc^{ac75REP}-*ph* at an MOI of 5 were fixed at the indicated time points. The cells were incubated with an anti-AC75 antibody (red). The nuclei were stained with Hoechst 33258 (blue).

levels in vAc^{ac75REP}-*ph*-infected cells following CHX treatment (**Figure 8C**). Accordingly, these data are consistent with the flow cytometry results presented above. Taken together, these data suggest that the AC75-F54S mutation affects the relative stability of AC75 in infected cells.

DISCUSSION

ac75 is a highly conserved gene in all sequenced baculovirus genomes except CuniNPV. According to previous reports, *ac75* is essential for nuclear egress and intranuclear microvesicle formation (Guo et al., 2017; Shi et al., 2017). Deletion of *ac75* resulted in a complete loss of viral growth in cultured cells, which precluded the investigation of AC75 function during late infection. In this study, we identified that Phe-54 was an important residue of AC75, and validated the important role of *ac75* in polyhedrin assembly and ODV embedding during OB formation.

Previous transcriptome analysis identified three typical late promoters located at 282, 93, and 14 nt upstream of the *ac75* start codon in *Trichoplusia ni* (*T. ni*) cells infected with AcMNPV (Chen et al., 2013). In this study, the transcription pattern revealed that *ac75* was transcribed from 12 h p.i. and this transcription persisted up to 72 h p.i. These observations are in accord with previous results showing that *ac75* was expressed at relatively high levels from 18 to 72 h p.i. in the AcMNPV-infected midgut of *T. ni* (Shrestha et al., 2018). In addition, western blot analysis revealed that AC75 was expressed during the late infection phase. Thus, *ac75* may play a role during the late stage of infection. Shi et al. (2017) and Guo et al. (2017) found that *ac75* was essential for nuclear egress and intranuclear microvesicle formation. However, the possible functions of *ac75*

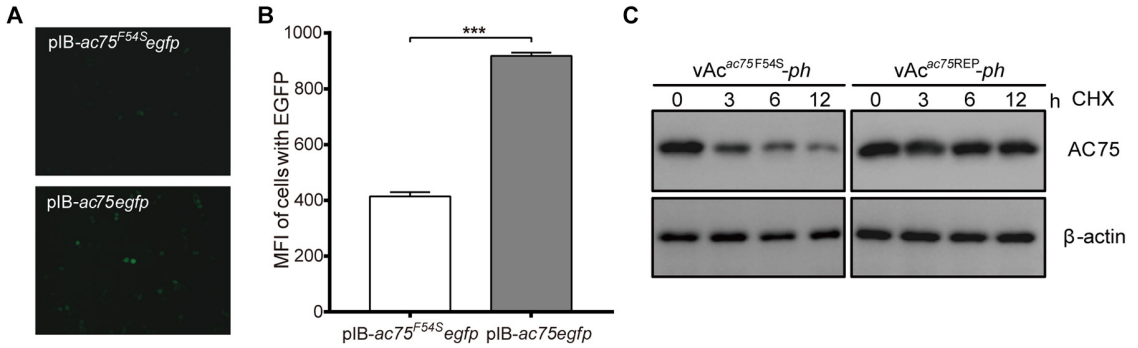


FIGURE 8 | Stability analysis of AC75. **(A)** Fluorescence analyses. Sf9 cells transfected with pIB-*ac75*^{F54S}*egfp* or pIB-*ac75egfp* plasmids were imaged at 48 h p.t. **(B)** Flow cytometry analyses. The mean fluorescence intensity of cells expressing the EGFP gene were analyzed by flow cytometry analysis. The values represent averages from three independent assays and error bars represent SD (*** indicates $P < 0.001$). **(C)** Western blot analysis. Sf9 cells infected with vAc^{ac75F54S}-*ph* or vAc^{ac75REP}-*ph* in triplicate at an MOI of 5 were treated with CHX at 24 h p.i. Lysates prepared at the indicated times thereafter were subjected to immunoblot analysis by using anti-AC75 and anti-actin antibodies.

in other processes during the late infection phase have not been investigated in detail.

Sequence alignment of AC75 homologs showed that Phe-54 and Gln-81 were completely conserved among the selected homologs. Phe is a hydrophobic amino acid with an aromatic side chain, whereas Gln has a polar, non-charged side chain. In this study, Phe-54 was mutated to glycine (F54G), alanine (F54A), or serine (F54S). These mutant viruses were transfected/infected into Sf9 cells, revealing that the AC75-F54S mutation produced 10-fold fewer progeny BVs than the wild-type virus, whereas the AC75-F54G mutation exhibited a modest two-fold reduction in viral titer when compared with that of the wild-type virus. The AC75-F54A mutation did not affect BV production (data not shown). Mutation of Gln-81 to glycine (Q81G), alanine (Q81A) or glutamic acid (Q81E) did not affect viral propagation (data not shown), indicating that residue Gln-81 may not be important for the functions of AC75. Thus, Phe-54 may be an important residue of AC75, and the AC75-F54S mutant virus was chosen for subsequent research.

In this study, novel functions of AC75 were investigated through characterization of the AC75-F54S partial loss-of-function mutant virus during the late infection phase. Electron microscopy revealed that the OBs of vAc^{ac75F54S}-*ph* had a ragged surface and contained markedly low numbers of ODVs and most OBs of the vAc^{ac75F54S}-*ph* were empty. In addition, the AC75-F54S mutation did not affect expression and localization of the polyhedrin. Thus, *ac75* may play a key role in the assembly of the polyhedrin and the embedding of ODVs during OB morphogenesis. The process of OB occlusion requires a complex integration of events, including envelopment of ODVs, polyhedrin assembly and incorporation of ODVs. The mature ODVs associate with dense concentrations of the polyhedral protein, which subsequently crystallizes around one or many ODVs to form OBs. The mechanisms that control crystallization at the ODV-polyhedrin interface are unresolved and the viral proteins required for this process are unknown. The surface layer (calyx or polyhedral envelope) of mature OBs is composed of carbohydrates and PEP (Whitt and Manning, 1988). The formation of this layer structure requires a functional P10 and is associated with nuclear fibrillar structures (Carpentier et al., 2008). The OBs of a *pep* or *p10* mutant virus are irregular and fragile with no outer layer (Vlak et al., 1988; Li et al., 2015). According to recent reports, the OB morphology was also abnormal in a P33 point mutation and a VP91 truncated mutation (Kuang et al., 2017; Zhou et al., 2019), suggesting that both P33 and VP91 are involved in OBs morphogenesis, which involves assembly of the polyhedrin and the embedding of ODVs. Therefore, *ac75* is the third gene identified to be involved with the assembly of the polyhedrin and embedding of ODVs. Further studies examining genes associated with OB formation will aid our understanding of this molecular mechanism.

Further investigation revealed that the AC75-F54S mutation reduced the amount of AC75, but did not affect localization in infected cells. We hypothesized that the lower amount of AC75 in vAc^{ac75F54S}-*ph*-infected cells affected the efficiency of nuclear egress and therefore resulted in lower amounts of BV production. Flow cytometry and western blot analyses revealed that the lower levels of AC75-F54S might be a result of impaired stability caused by AC75-F54S mutation. The loss of persistence of AC75 may be caused by structural disorder. Thus, the AC75-F54S mutation may have caused mis-folding of AC75, which further affected its proper functions in the AcMNPV life cycle. Moreover, incorrectly folded AC75 may be degraded by the 20S proteasome (Asher et al., 2006; Tompa et al., 2008). Further analysis is required to determine the exact functional role of Phe-54 in AC75 activity in AcMNPV-infected cells.

In conclusion, we revealed that the residue Phe-54 in AC75 is important for the proper folding of this protein, and AC75 is involved in polyhedrin assembly and ODV embedding during OB formation. Thus, *ac75* is a versatile gene associated with nuclear egress of nucleocapsids and OB morphogenesis. Moreover, our results provide evidence for a pivotal role of *ac75* in BV production and OB formation, and further studies of AC75 are required to elucidate the relation between BV formation and OB morphogenesis. These data may help uncover the mechanism of OB morphogenesis of baculoviruses.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/ **Supplementary Material**.

ETHICS STATEMENT

The animal study was reviewed and approved by the Institutional Review Board, Wuhan Institute of Virology, Chinese Academy of Sciences.

AUTHOR CONTRIBUTIONS

XC, XS, and JH designed the experiments and wrote the manuscript. XC, XY, and CL carried out the experiments and analyzed the data. XC, JY, XS, and JH checked and finalized the manuscript. All authors contributed to manuscript revision, read, and approved the submitted version.

FUNDING

This work was supported by the National Key Research and Development Program of China (2018YFE0121900 to JH) and the WIV "One-Three-Five" strategic program (Y602111SA1 to XS). The funders had no role in study design, data collection, and interpretation, or the decision to submit the work for publication.

ACKNOWLEDGMENTS

We thank the Core Facility and Technical Support of Wuhan Institute of Virology, CAS, for the technical assistance in antibody preparation (Youling Zhu), fluorescence microscopy (Ding Gao and Juan Min) and electron microscope (Pei Zhang, Anna Du, and Bichao Xu).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2021.663506/full#supplementary-material

Supplementary Figure 1 | Sequence alignment of the AC75 homologs from 54 alphabaculoviruses and 12 betabaculoviruses. Amino acid sequences were aligned using Clustal X-2.0. Black shading denotes identical conserved residues, whereas gray shading indicates less conserved sites. Residues Phe-54 and Gln-81 were absolutely conserved among the selected homologs. The GenBank accession numbers of the amino acid sequences are presented in **Supplementary Table 1**.

Supplementary Figure 2 | Immunoelectron microscopy analysis. Sf9 cells were infected with vAc-*ph*, vAc^{ac75F54S}-*ph,* or vAc^{ac75REP}-*ph* at an MOI of five and harvested at 48 h p.i. Ultrathin sections were probed with the anti-AC75 antibody as the primary antibody and goat anti-rabbit IgG coated with gold particles (10 nm) as the secondary antibody.

Supplementary Figure 3 | Immunofluorescence analysis of localization and expression of AC75 in vAc-*ph*-infected cells. Sf9 cells were infected with vAc-*ph* (MOI = 5) and fixed at the indicated time points. The cells were incubated with an anti-AC75 antibody (red). The nuclei were stained with Hoechst 33258 (blue).

Supplementary Table 1 | The GenBank accession numbers of the amino acid sequences of AC75 homolog proteins selected in this study.

REFERENCES

Asher, G., Reuven, N., and Shaul, Y. (2006). 20S proteasomes and protein degradation "by default". *Bioessays* 28, 844–849. doi: 10.1002/bies.20447

Ayres, M. D., Howard, S. C., Kuzio, J., Lopez-Ferber, M., and Possee, R. D. (1994). The complete DNA sequence of Autographa californica nuclear polyhedrosis virus. *Virology* 202, 586–605. doi: 10.1006/viro.1994.1380

Biswas, S., Willis, L. G., Fang, M., Nie, Y., and Theilmann, D. A. (2017). Autographa californica Nucleopolyhedrovirus AC141 (Exon0), a Potential E3 Ubiquitin Ligase, Interacts with Viral Ubiquitin and AC66 To Facilitate Nucleocapsid Egress. *J. Virol.* 92, 01713–01717e.

Blissard, G. W., and Theilmann, D. A. (2018). Baculovirus Entry and Egress from Insect Cells. *Annu. Rev. Virol.* 5, 113–139. doi: 10.1146/annurev-virology-092917-043356

Blissard, G. W., and Wenz, J. R. (1992). Baculovirus gp64 envelope glycoprotein is sufficient to mediate pH-dependent membrane fusion. *J. Virol.* 66:6829.

Braunagel, S. C., and Summers, M. D. (2007). Molecular biology of the baculovirus occlusion-derived virus envelope. *Curr. Drug Targets* 8, 1084–1095.

Byers, N. M., Vandergaast, R. L., and Friesen, P. D. (2016). Baculovirus Inhibitor-of-Apoptosis Op-IAP3 Blocks Apoptosis by Interaction with and Stabilization of a Host Insect Cellular IAP. *J. Virol.* 90, 533–544. doi: 10.1128/JVI.02320-15

Carpentier, D. C. J., Griffiths, C. M., and King, L. A. (2008). The baculovirus P10 protein of Autographa californica nucleopolyhedrovirus forms two distinct cytoskeletal-like structures and associates with polyhedral occlusion bodies during infection. *Virology* 371, 278–291.

Chen, L., Hu, X., Xiang, X., Yu, S., Yang, R., and Wu, X. (2012). Autographa californica multiple nucleopolyhedrovirus odv-e25 (Ac94) is required for budded virus infectivity and occlusion-derived virus formation. *Arch. Virol.* 157, 617–625. doi: 10.1007/s00705-011-1211-9

Chen, X., Yang, X., Lei, C., Qin, F., Sun, X., and Hu, J. (2020). Autographa californica Multiple Nucleopolyhedrovirus orf13 is Required for Efficient Nuclear Egress of Nucleocapsids. *Virol. Sin*. Preprint. doi: 10.1007/s12250-021-00353-3

Chen, Y. R., Zhong, S., Fei, Z., Hashimoto, Y., Xiang, J. Z., Zhang, S., et al. (2013). The transcriptome of the baculovirus Autographa californica multiple nucleopolyhedrovirus in Trichoplusia ni cells. *J. Virol.* 87, 6391–6405. doi: 10.1128/JVI.00194-13

Datsenko, K. A., and Wanner, B. L. (2000). One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proc. Natl. Acad. Sci. U S A.* 97, 6640–6645. doi: 10.1073/pnas.120163297

Fang, M., Dai, D., and Theilmann, D. A. (2007). Autographa californica multiple nucleopolyhedrovirus EXON0 (ORF141) is required for efficient egress of nucleocapsids from the nucleus. *J. Virol.* 81, 9859–9869.

Gross, C. H., Russell, R. L., and Rohrmann, G. F. (1994). Orgyia pseudotsugata baculovirus p10 and polyhedron envelope protein genes: analysis of their relative expression levels and role in polyhedron structure. *J. Gen. Virol.* 75, 1115–1123. doi: 10.1099/0022-1317-75-5-1115

Guo, Y. J., Fu, S. H., and Li, L. L. (2017). Autographa californica multiple nucleopolyhedrovirus *ac75* is required for egress of nucleocapsids from the nucleus and formation of de novo intranuclear membrane microvesicles. *PLoS One* 12:e0185630. doi: 10.1371/journal.pone.0185630

Hepp, S. E., Borgo, G. M., Ticau, S., Ohkawa, T., and Welch, M. D. (2018). Baculovirus AC102 Is a Nucleocapsid Protein That Is Crucial for Nuclear Actin Polymerization and Nucleocapsid Morphogenesis. *J. Virol.* 92, 111–118e. doi: 10.1128/JVI.00111-18

Hodgson, J. J., Arif, B. M., and Krell, P. J. (2007). Reprogramming the chiA expression profile of Autographa californica multiple nucleopolyhedrovirus. *J. Gen. Virol.* 88(Pt 9), 2479–2487. doi: 10.1099/vir.0.82863-0

Hu, Z., Yuan, M., Wu, W., Liu, C., Yang, K., and Pang, Y. (2010). Autographa californica multiple nucleopolyhedrovirus ac76 is involved in intranuclear microvesicle formation. *J. Virol.* 84, 7437–7447. doi: 10.1128/JVI.02103-09

Ji, X., Zhang, C., Fang, Y., Zhang, Q., Lin, L., Tang, B., et al. (2015). Isolation and characterization of glacier VMY22, a novel lytic cold-active bacteriophage of Bacillus cereus. *Virol. Sin.* 30, 52–58. doi: 10.1007/s12250-014-3529-4

Jones, P., Binns, D., Chang, H. Y., Fraser, M., Li, W., McAnulla, C., et al. (2014). InterProScan 5: genome-scale protein function classification. *Bioinformatics* 30, 1236–1240. doi: 10.1093/bioinformatics/btu031

Ke, J., Wang, J., Deng, R., and Wang, X. (2008). Autographa californica multiple nucleopolyhedrovirus *ac66* is required for the efficient egress of nucleocapsids from the nucleus, general synthesis of preoccluded virions and occlusion body formation. *Virology* 374, 421–431. doi: 10.1016/j.virol.2007.12.033

Kuang, W., Zhang, H., Wang, M., Zhou, N. Y., Deng, F., Wang, H., et al. (2017). Three Conserved Regions in Baculovirus Sulfhydryl Oxidase P33 Are Critical for Enzymatic Activity and Function. *J. Virol.* 91, 1158–1117e. doi: 10.1128/JVI.01158-17

Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., McGettigan, P. A., McWilliam, H., et al. (2007). Clustal W and Clustal X version 2.0. *Bioinformatics* 23, 2947–2948. doi: 10.1093/bioinformatics/btm404

Li, C., Wen, A., Shen, B., Lu, J., Huang, Y., and Chang, Y. (2011). FastCloning: a highly simplified, purification-free, sequence- and ligation-independent PCR cloning method. *BMC Biotechnol.* 11:92. doi: 10.1186/1472-6750-11-92

Li, J., Zhou, Y., Lei, C., Fang, W., and Sun, X. (2015). Improvement in the UV resistance of baculoviruses by displaying nano-zinc oxide-binding peptides on the surfaces of their occlusion bodies. *Appl. Microbiol. Biotechnol.* 99, 6841–6853. doi: 10.1007/s00253-015-6581-6

Li, Y., Shen, S., Hu, L., Deng, F., Vlak, J. M., Hu, Z., et al. (2018). The functional oligomeric state of tegument protein GP41 is essential for baculovirus BV and ODV assembly. *J. Virol.* 92, 2083–2017e.

Marchler-Bauer, A., Bo, Y., Han, L., He, J., Lanczycki, C. J., Lu, S., et al. (2017). CDD/SPARCLE: functional classification of proteins via subfamily domain architectures. *Nucleic Acids Res.* 45, D200–D203. doi: 10.1093/nar/gkw1129

McCarthy, C. B., Dai, X., Donly, C., and Theilmann, D. A. (2008). Autographa californica multiple nucleopolyhedrovirus *ac142*, a core gene that is essential for BV production and ODV envelopment. *Virology* 372, 325–339. doi: 10.1016/j.virol.2007.10.019

Ohkawa, T., and Welch, M. D. (2018). Baculovirus Actin-Based Motility Drives Nuclear Envelope Disruption and Nuclear Egress. *Curr. Biol.* 28, 2153–2159. doi: 10.1016/j.cub.2018.05.027

Qin, F., Xu, C., Hu, J., Lei, C., Zheng, Z., Peng, K., et al. (2019). Dissecting the Cell Entry Pathway of Baculovirus by Single-Particle Tracking and Quantitative Electron Microscopic Analysis. *J. Virol.* 93, 33–19e. doi: 10.1128/JVI.00033-19

Qiu, J., Tang, Z., Cai, Y., Wu, W., Yuan, M., and Yang, K. (2019). The Autographa californica Multiple Nucleopolyhedrovirus *ac51* Gene Is Required for Efficient Nuclear Egress of Nucleocapsids and Is Essential for In Vivo Virulence. *J. Virol.* 93, 1923–1918e. doi: 10.1128/JVI.01923-18

Rohrmann, G. F. (2019). *Baculovirus Molecular Biology*, 4th Edn. Bethesda, MD: National Center for Biotechnology Information.

Russell, R. L. Q., and Rohrmann, G. F. (1990). A baculovirus polyhedron envelope protein: Immunogold localization in infected cells and mature polyhedra. *Virology* 174, 177–184.

Shi, A., Hu, Z., Zuo, Y., Wang, Y., Wu, W., Yuan, M., et al. (2017). Autographa californica Multiple Nucleopolyhedrovirus *ac75* Is Required for the Nuclear Egress of Nucleocapsids and Intranuclear Microvesicle Formation. *J. Virol.* 92, 01509–01517e.

Shorey, H. H., and Hale, R. L. (1965). Mass-Rearing of the Larvae of Nine Noctuid Species on a Simple Artificial Medium. *J. Econ. Entomol.* 58, 522–524.

Shrestha, A., Bao, K., Chen, Y. R., Chen, W., Wang, P., Fei, Z., et al. (2018). Global Analysis of Baculovirus Autographa californica Multiple Nucleopolyhedrovirus Gene Expression in the Midgut of the Lepidopteran Host Trichoplusia ni. *J. Virol.* 92, 1277–1218e. doi: 10.1128/JVI.01277-18

Tao, X. Y., Choi, J. Y., Kim, W. J., An, S. B., Liu, Q., Kim, S. E., et al. (2015). Autographa californica multiple nucleopolyhedrovirus ORF11 is essential for budded-virus production and occlusion-derived-virus envelopment. *J. Virol.* 89, 373–383. doi: 10.1128/JVI.01742-14

Tao, X. Y., Choi, J. Y., Kim, W. J., Lee, J. H., Liu, Q., Kim, S. E., et al. (2013). The Autographa californica multiple nucleopolyhedrovirus ORF78 is essential for budded virus production and general occlusion body formation. *J. Virol.* 87, 8441–8450. doi: 10.1128/JVI.01290-13

Tompa, P., Prilusky, J., Silman, I., and Sussman, J. L. (2008). Structural disorder serves as a weak signal for intracellular protein degradation. *Proteins* 71, 903–909. doi: 10.1002/prot.21773

van Oers, M. M., and Vlak, J. M. (2007). Baculovirus genomics. *Curr. Drug Targets* 8, 1051–1068.

Vlak, J. M., Klinkenberg, A. F., Zaal, K. J. M., Usmany, M., Klinge-Roode, E. C., Geervliet, J. B. F., et al. (1988). Functional Studies on the p10 Gene of Autographa californica Nuclear Polyhedrosis Virus Using a Recombinant Expressing a p10-β-Galactosidase Fusion Gene. *J. General Virol.* 69(Pt 4), 765.

Wang, Y., Cai, Q., Chen, J., Huang, Z., Wu, W., Yuan, M., et al. (2019). Autographa Californica Multiple Nucleopolyhedrovirus P48 (Ac103) Is Required for the Efficient Formation of Virus-Induced Intranuclear Microvesicles. *Virol. Sin.* 34, 712–721. doi: 10.1007/s12250-019-00147-8

Whitt, M. A., and Manning, J. S. (1988). A phosphorylated 34-kDa protein and a subpopulation of polyhedrin are thiol linked to the carbohydrate layer surrounding a baculovirus occlusion body. *Virology* 163, 33–42.

Williams, G. V., Rohel, D. Z., Kuzio, J., and Faulkner, P. (1989). A cytopathological investigation of Autographa californica nuclear polyhedrosis virus p10 gene function using insertion/deletion mutants. *J. General Virol.* 70(Pt 1), 187.

Xu, C., Wang, J., Yang, J., Lei, C., Hu, J., and Sun, X. (2019). NSP2 forms viroplasms during Dendrolimus punctatus cypovirus infection. *Virology* 533, 68–76. doi: 10.1016/j.virol.2019.05.005

Young, J. C., MacKinnon, E. A., and Faulkner, P. (1993). The Architecture of the Virogenic Stroma in Isolated Nuclei of Spodoptera frugiperda Cells in Vitro Infected by Autographa californica Nuclear Polyhedrosis Virus. *J. Struct. Biol.* 110, 141–153. doi: 10.1006/jsbi.1993.1015

Yuan, M., Huang, Z., Wei, D., Hu, Z., Yang, K., and Pang, Y. (2011). Identification of Autographa californica nucleopolyhedrovirus *ac93* as a core gene and its requirement for intranuclear microvesicle formation and nuclear egress of nucleocapsids. *J. Virol.* 85, 11664–11674.

Yuan, M., Wu, W., Liu, C., Wang, Y., Hu, Z., Yang, K., et al. (2008). A highly conserved baculovirus gene *p48* (*ac103*) is essential for BV production and ODV envelopment. *Virology* 379, 87–96. doi: 10.1016/j.virol.2008.06.015

Yue, Q., Yu, Q., Yang, Q., Xu, Y., Guo, Y., Blissard, G. W., et al. (2018). Distinct Roles of Cellular ESCRT-I and ESCRT-III Proteins in Efficient Entry and Egress of Budded Virions of Autographa californica Multiple Nucleopolyhedrovirus. *J. Virol.* 92, 1636–1617e. doi: 10.1128/JVI.01636-17

Zhou, F., Kuang, W., Wang, X., Hou, D., Huang, H., Sun, X., et al. (2019). The cysteine-rich region of a baculovirus VP91 protein contributes to the morphogenesis of occlusion bodies. *Virology* 535, 144–153. doi: 10.1016/j.virol.2019.06.016

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Microbiology

REVIEW
published: 14 April 2021
doi: 10.3389/fmicb.2021.649690

Congenital Human Cytomegalovirus Infection Inducing Sensorineural Hearing Loss

*Wenwen Xia[1], Hui Yan[1], Yiyuan Zhang[2], Congcong Wang[3], Wei Gao[4], Changning Lv[1], Wentao Wang[1] and Zhijun Liu[3]**

[1]School of Clinical Medicine, Weifang Medical University, Weifang, China, [2]Key Laboratory of Carcinogenesis and Translational Research (Ministry of Education), Gastrointestinal Cancer Center, Peking University Cancer Hospital and Institute, Beijing, China, [3]Department of Microbiology, Weifang Medical University, Weifang, China, [4]Key Lab for Immunology in Universities of Shandong Province, School of Clinical Medicine, Weifang Medical University, Weifang, China

Human cytomegalovirus (HCMV) is the primary cause of congenital infections. Despite its clinical significance, congenital HCMV infection is frequently overlooked clinically since most affected infants are asymptomatic. Sensorineural hearing loss (SNHL) is one of the most widely known disorders caused by congenital HCMV infection. The potential mechanism, however, remains unknown to date. The mechanism by which congenital HCMV infection induces sensorineural deafness has been partly characterized, leading to advancements in diagnosis, therapy, and prevention strategies. HCMV-induced hearing loss primarily involves immune responses, the release of inflammatory factors by natural killer (NK) cells, apoptosis of cochlear spiral ganglion, and potential changes due to vascular dysfunction. The diagnosis of HCMV induced SNHL includes serological examination to mothers, imaging, and amniotic fluid examination. Ganciclovir, mainly used for antiviral therapy and behavioral prevention, can, to some degree, prevent congenital HCMV infection. The role of HCMV infection in hearing loss needs further investigation since the mechanism of hearing loss caused by cytomegalovirus infection is not well understood. Although some advancement has been made in diagnosing and treating SNHL, more improvement is needed. A comprehensive understanding of cytomegalovirus's pathogenesis is of key importance for preventing, diagnosing, and treating SNHL.

Keywords: cytomegalovirus, hearing loss, mechanism, diagnosis, development

OPEN ACCESS

Edited by:
Julien Andreani,
IHU Mediterranee Infection, France

Reviewed by:
Tatsuo Suzutani,
Fukushima Medical University, Japan
Valentina Dell'Oste,
University of Turin, Italy

****Correspondence:***
Zhijun Liu
zhijun.liu@wfmc.edu.cn

Specialty section:
This article was submitted to Virology, a section of the journal Frontiers in Microbiology

Received: *05 January 2021*
Accepted: *22 March 2021*
Published: *14 April 2021*

Citation:
Xia W, Yan H, Zhang Y, Wang C, Gao W, Lv C, Wang W and Liu Z (2021) Congenital Human Cytomegalovirus Infection Inducing Sensorineural Hearing Loss. Front. Microbiol. 12:649690. doi: 10.3389/fmicb.2021.649690

INTRODUCTION

Human cytomegalovirus (HCMV) belongs to the herpesviridae family, transmitted from mother-to-child *in utero*, intrapartum, and *via* breastfeeding (Davis et al., 2017; Dobbie, 2017). The infection's major routes are sexual transmission and contact with body fluids, such as semen, cervical or vaginal secretions, urine, and blood. HCMV can also be transmitted through saliva and breast milk after birth. Congenital HCMV infection is one of the most common congenital infections. HCMV infection can lead to an asymptomatic clinical situation as well as to a severely symptomatic patient. The clinical symptoms include petechiae, microcephaly, chorioretinitis, hepatosplenomegaly, and growth retardation. Many congenital

HCMV infections have long-term sequelae, and sensorineural hearing loss (SNHL) is the most common sequela (Bartlett et al., 2017). At present, however, the mechanism of HCMV induced SNHL remains unclear. Also, there is no effective way to prevent HCMV transmission from mother to infant or effective treatment. In this review, we summarized recent studies on hearing loss associated with HCMV. The possible mechanisms of HCMV induced hearing loss and the progress in the diagnosis and treatment of hearing loss were reviewed and discussed.

Possible Mechanism of HCMV Infection Developing to SNHL

Congenital HCMV infection is considered the most common non-genetic cause of SNHL (Dhondt et al., 2019). We have learned the current research progress of SNHL associated with HCMV by gathering many pieces of literature (**Table 1**). HCMV can invade different parts of the auditory pathway, such as the inner ear, the middle ear, the afferent, and efferent nerve fibers: the damage after the infection of these compartments ultimately leads to deafness. In the early stage of viremia, virus particles can directly enter the inner ear from the blood (the most significant infection pathway) or through the cochlea's aqueduct from the subarachnoid cavity. HCMV can also invade the inner ear through the round window membrane of the middle ear. Even though the infection can spread to the other compartments of the ear and the auditory nerve, the inner ear remains the most significant section in terms of the pathogenesis of deafness. In particular, HCMV causes microcirculation disorders, tissue hyperplasia in the organs of Corti, and cellular damage that leads to spiral ganglion neurons (SGN) cells' apoptosis and changes of the endocochlear potential (EP). Recent studies have revealed that part of the damage-causing deafness also involves the immune response induced by HCMV infection: the activity of NK cells and the expression of proinflammatory cytokines lead to the destruction of the blood-labyrinth barrier (Li et al., 2014; Bradford et al., 2015; Almishaal et al., 2017). Damage to the hair cells may cause hearing loss, as some investigators observed injury to the outer hair cells in a mouse cytomegalovirus (CMV) congenital infection model. This hearing disorder can be treated with a cochlear implant, indicating that the nerve is intact.

HCMV Infection Induces the Inflammatory Responses

The inflammation may be involved in SNHL (Tian et al., 2018). Researchers established murine HCMV infection models of neonatal mice to study the mechanism of SNHL and tried to infer the mechanism of SNHL caused by HCMV infection by studying the infection model of mice. Bradford et al. found that inflammation was the important mechanism of hearing loss in mice (Bradford et al., 2015). Zhuang et al. (2018) suggested that the hearing loss caused by HCMV infection was associated with reactive oxygen species (ROS) induced inflammation. HCMV increases ROS level and activates nucleotide-binding oligomerization domain-like receptor protein 3 (NLRP3) inflammatory bodies in the cochlea and cultured SGN, resulting in the activation of Caspase 1 and increasing the maturation and release of IL-1 beta and IL-18 (Shi et al., 2017; **Figure 1**). Similarly, Bradford et al. (2015) established a viral infection model of sensorineural deafness induced by a viral infection in newborn mice, and their findings revealed the density reduction of SGN and persistent inflammation in the cochlear tissues of the deaf mice, indicating that inflammation is an important part of the mechanism of hearing loss induced by HCMV infection.

TABLE 1 | Possible mechanisms of HCMV related SNHL.

Main points of view	Year	Researchers
Immune responses		
Activating inflammatory responses, increasing ROS, and activating NLRP3 inflammatory cells, causing Caspase 1 activation and increasing the maturation and release of IL-1 beta and IL-18.	2018	Zhuang, W., et al.
The role of the interaction between the M157 on the virus surface and the LY49 cell surface receptor on the NK cells in HCMV related hearing loss.	2018	Almishaal, A. A., et al.
Destroying the integrity of BLB, leading to the destruction of microcirculation and the homeostasis of the internal environment.	2014	Li, X., et al.
Degeneration and injury of cells		
Cell apoptosis	2013	Schmutzhard, J., et al.
Migration to the inner ear, hearing impairment is associated with the poor maintenance of the EP caused by strial dysfunction.	2017	Carraro, M., et al.

HCMV Infection Induces the Immune Response of NK Cells

Natural killer (NK) cells play a role in the early immune response to viruses. NK cells participate in the immune response by expressing cytokines and antibody-dependent cell-mediated cytotoxicity (ADCC; Hammer et al., 2018). Here, the interaction between the M157 (encoded by the virus) and the LY49H cell surface receptor of the NK cells in murine HCMV-related hearing loss will be mainly introduced (Almishaal et al., 2017). At the early stage of infection, the LY49H receptor recognizes M157, and the interaction between M157 and Ly49 H receptor triggers the activation of NK cells and the elimination of the infected cells (Almishaal et al., 2017). This interaction may help NK cells recognize murine HCMV infection early and prevent SGN cell apoptosis and hearing loss. Currently, however, there is no evidence showing that interfering in NK cells can improve hearing. Thus, NK cells have not been used as therapeutic targets.

Abbreviations: HCMV, Human cytomegalovirus; CMV, Cytomegalovirus; SNHL, Sensorineural hearing loss; SGN, Spiral ganglion neurons; EP, Endocochlear potential; ROS, Reactive oxygen species; NLRP3, Nucleotide-binding oligomerization domain-like receptor protein 3; NK, Natural killer; ADCC, Antibody-dependent cell-mediated cytotoxicity; SV, Striavascularis; BLB, Blood labyrinth barrier; DBS, Dry blood spot test; PCR, Polymerase chain reaction; HIG, Hyperimmune globulin.

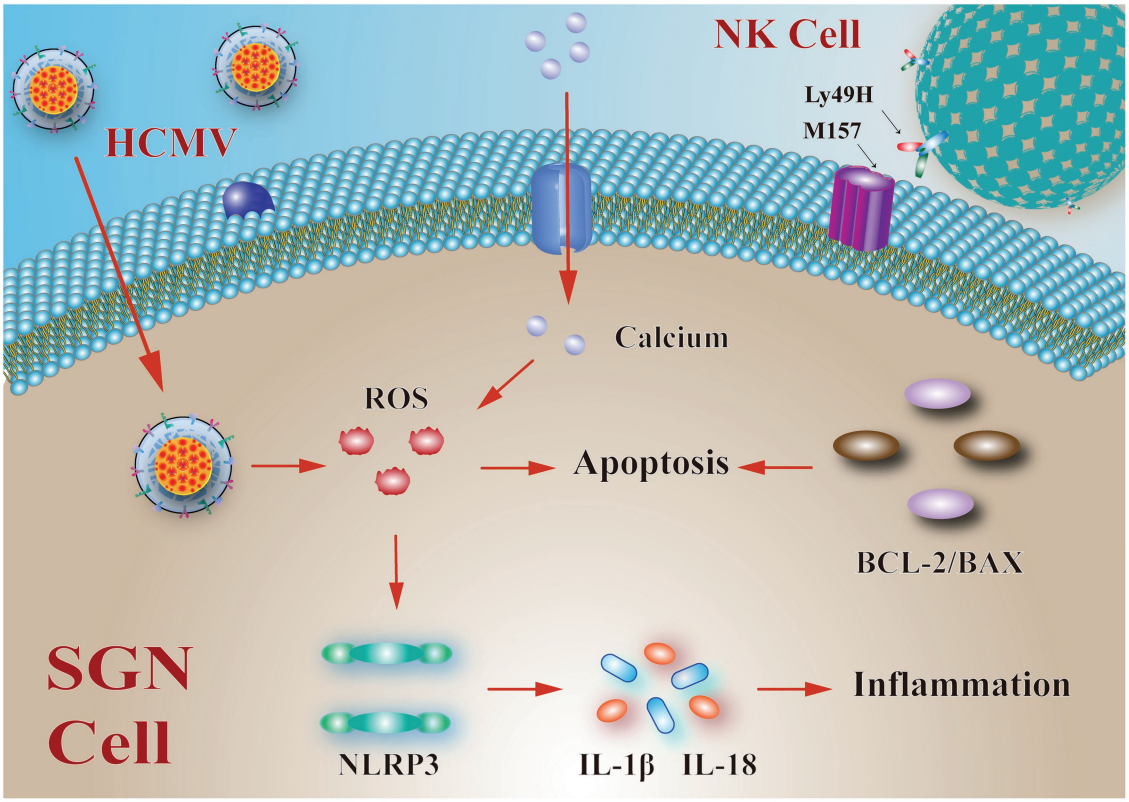


FIGURE 1 | Possible mechanisms and pathways of human cytomegalovirus (HCMV) infection in spiral ganglion neurons (SGN) cells. The mechanism of sensorineural hearing loss (SNHL) may be related to the apoptosis of SGN cells. The HCMV infection can increase reactive oxygen species (ROS) levels, activate nucleotide-binding oligomerization domain-like receptor protein 3 (NLRP3) inflammatory bodies in cochlear and SGN, and activate Caspase 1 to increase the maturation and release of IL-1 beta and IL-18, thus leading to inflammatory responses. Continuous increase in Ca_2^+ also causes an increase in ROS. Bcl-2 and Bax are also involved in the apoptosis of SGN cells. Also, the interaction between LY49H on the surface of NK cells and M157 expressed by HCMV is associated with the apoptosis of SGN.

HCMV Infection Might Destroy the Integrity of the Blood-Labyrinth Barrier

The normal physiological function of auditory pathways, based on the normal function of microcirculation and blood labyrinth barrier (BLB), is the basis for the normal conduction of auditory signals. The cochlear BLB, located in the stria vascularis (SV), plays an important role in maintaining the homeostasis of the cochlea, preventing toxic substances from flowing into the inner ear and selectively transferring ions, fluids, and nutrients to the cochlea (Juhn and Rybak, 1981; Wu et al., 2017).

Li et al. (2014) also established a CMV infection model in newborn mice and found that the BLB permeability of the CMV infection group was much higher. They hypothesized that CMV infection might destroy the integrity of BLB, which further leads to the destruction of microcirculation and the homeostasis of the internal environment. In addition, monocytes infected with HCMV may disseminate through blood circulation, resulting in systemic infection and further disrupting the integrity of BLB (Firbas et al., 1981; Kimura et al., 1990; Li et al., 2014).

HCMV Affects the Apoptosis of SGN Cells

Human cytomegalovirus infection is a complex process. Besides the host immune reaction, it also involves virus reactivation, cell damage, and cell apoptosis (Li et al., 2016). Apoptosis of SGN cells may be the primary cause of hearing impairment (Schmutzhard et al., 2013). SGN cells play an important role in the transmission of electrical signals. The SGNs are referred to as the first level of neurons of the auditory system, and during the auditory signal transmission, SGN receives electrical signal input from cochlear hair cells and transmits it from the cochlea to the cochlear nucleus; subsequently, the electrical signals are transmitted to the auditory cortex (Bailey and Green, 2014). The dysfunction of SGN caused by HCMV can often lead to SNHL. However, the specific mechanisms of SGN apoptosis are not fully understood (Li et al., 2016).

Two mechanisms may be associated with the apoptosis of SGN cells. One mechanism is through the increase of Ca_2^+. It has been reported that a persistent increase in Ca_2^+ could lead to SGN apoptosis after HCMV infection (Li et al., 2016). The other mechanism may be related to Bax and Bcl-2 protein ratio. It is found that SGN apoptosis may be mediated by Bcl-2 and Bax (Li et al., 2016). Bcl-2 family proteins are the major regulators of apoptosis. In particular, Bcl-2 protein could inhibit apoptosis, while Bax could promote it. In a word, after HCMV infection, the Bax level increased, Bcl-2 and Bcl-2/Bax ratio decreased in cells, indicating that HCMV infection can induce apoptosis.

HCMV Infection Leads to the Lesion of Stria Vascularis and Causes the Poor Development of EP

In the mouse model and by using a modified corrosion casting technique, Carraro et al. (2017) found that after HCMV migration to the inner ear through the blood or cerebrospinal fluid, the stria vascular was the primary site for HCMV infection, and they believed that hearing impairment might be associated with the poor maintenance of the EP caused by strial dysfunction. The HCMV infection may lead to the damage of the Stria Vascularis (the main target of HCMV in the inner ear), destroying the potassium cycle and poor development of the pool of auditory sensory cells EP (Teissier et al., 2016).

The EP is essential for hearing, and it provides approximately half of the driving force for the transduction current in auditory hair cells (Cohen-Salmon et al., 2007). It is found that the maintenance of EP is related to the high potassium. The high potassium depends on channel proteins and transporters in the inner ear, ensuring that potassium ions circulate continuously into the endolymph (Mittal et al., 2017). The EP is produced by the SV, and SV is the prime target of HCMV (Cohen-Salmon et al., 2007; Teissier et al., 2016). HCMV infection can damage SV, which can alter the potassium cycle. The potassium ion cycle change reduces the inner ear lymphatic potential required by depolarization of inner ear sensory cells; consequently, SNHL may occur (Gabrielli et al., 2013).

It has been reported that the interaction between HCMV and connexin26 plays an important role in potassium circulation in the inner ear (Teissier et al., 2016). It is hypothesized that the mutation of connexin26 can cause a deficiency of potassium ion circulation. Although this hypothesis is widely mentioned, it has not been proved (Zhao, 2017).

DIAGNOSIS

Congenital HCMV diagnosis includes prenatal diagnosis and neonatal diagnosis. The first step in prenatal diagnosis of congenital HCMV is to determine the primary and secondary infection indicators of pregnant women by serological tests, and the second step is to determine whether the fetus is infected by noninvasive and invasive examination (Yinon et al., 2018). The diagnosis of neonates is achieved by virus detection within the first 3 weeks of life (Swanson and Schleiss, 2013; Lim and Lyall, 2017). Imaging and genetic testing had the highest yield in evaluating children with SNHL and were the most performed. HCMV testing was valuable in neonates that failed newborn hearing screening (Wentland et al., 2018).

Diagnosis of Maternal Infection

Serological tests can be used to diagnose primary HCMV infections. Primary maternal infections can be identified by serologic testing using IgG and IgM serology: IgG avidity testing will be used only when CMV-specific IgM antibodies are positive (Davis et al., 2017). Theoretically, only IgM indicates HCMV acute infection, a sensitive marker of primary HCMV infection. However, IgM is less specific and has a high false-positive rate because HCMV IgM is also produced during the virus's reactivation and still exists after some primary infection in some individuals. Therefore, IgM positive alone is not enough to diagnose HCMV primary infection (Prince and Lape-Nixon, 2014). Rarely the IgM response may be transient with CMV IgM antibodies lasting only a short time and maybe undetected even in the context of recent primary infection. However, in most primary infections, the IgM antibodies persist for weeks at levels that are unlikely to be missed by standard commercial kits. The IgG avidity essay is a tool that can be used to more accurately detect a primary infection than IgM alone (Navti et al., 2020).

Human cytomegalovirus IgG affinity is a sensitive and specific indicator to identify recent HCMV infection in pregnant women. IgG affinity is defined as the strength of IgG binding to an antigen epitope expressed by a given protein, which matures for 6 months after primary infection. Low HCMV IgG affinity is an accurate indicator of primary infection in the first 3 or 4 months of infection, while according to IgG affinity, the most recent primary infections can be excluded (Manicklal et al., 2013).

The most direct index of HCMV infection is serum transformation during pregnancy. However, this method is ineffective because of the lack of pre-pregnancy antibody screening procedures to select seronegative women. Diagnosis of HCMV infection is a complex process, but reliable diagnosis can be made by low IgG affinity and evidence of seroconversion (Swanson and Schleiss, 2013; Juckstock et al., 2015; Naing et al., 2016).

Diagnosis of Fetal Infection

Ultrasound examination of the fetus is the main noninvasive evaluation method in patients with suspected or confirmed HCMV infection (Naing et al., 2016). The ultrasonographic features of common fetal HCMV infection include high intestinal echo, hydronephrosis, fetal edema, hepatomegaly, periventricular echo density, ventricular expansion, cerebellum, and overall growth retardation (Juckstock et al., 2015). However, ultrasonography can only be used as an auxiliary examination because other intrauterine infections and fetal diseases also produce the same characteristics. These characteristics can only be observed in less than 25% of infected fetuses (Lipitz et al., 2010). Normal imaging cannot rule out the development of hearing loss and minor neurodevelopmental abnormalities (Lipitz et al., 2020).

An amniotic fluid examination is the first method for diagnosing an infected fetus (Fowler and Boppana, 2018). Amniocentesis could be used for HCMV virus culture and PCR to diagnose fetuses (Mestas, 2016). Since maternal infection and fetal infection are detectable at least 6–8 weeks after infection, amniocentesis should be performed at least 7 weeks after 20–21 weeks of pregnancy (Manicklal et al., 2013; Chiopris et al., 2020). Because the viral particles are excreted into the amniotic fluid through the fetus's urine, HCMV cannot be detected by amniocentesis in the amniotic fluid until the fetal kidney system is fully functioning. If the amniocentesis is carried out shortly after the early pregnancy or after the

mother infection diagnosis, the positive test results are reliable evidence for the fetal infection diagnosis. If the examination results are negative, a repeated examination is needed after the pregnancy (Fowler and Boppana, 2018).

PCR is more sensitive (70–90%) and practical to prenatal diagnosis than HCMV culture (Bodeus et al., 1999; Revello and Gerna, 2002). However, given the false-positive results caused by the PCR method, the combination of PCR and virus culture is more practically used (Enders et al., 2001; Gouarin et al., 2001; Revello et al., 2003).

Diagnosis of the HCMV-Infected Newborns

Most newborns with congenital HCMV infection are asymptomatic at birth, resulting in a low detection rate of HCMV infection. Approximately 15–25% of asymptomatic patients may have neurological sequelae, e.g., SNHL. Many patients with hearing loss are detected by neonatal hearing screening; however, regarding delayed SNHL, the effectiveness of newborn hearing screening tests for SNHL is limited. Neonatal hearing screening cannot predict the potential of hearing loss. Therefore, screening for HCMV infection is particularly important (Moteki et al., 2018). The principal method of clinical diagnosis of HCMV infection is serological detection and cultivation of the virus.

As samples of urine and saliva from newborns infected with HCMV are mostly rich in viruses, the isolation of virus from the urine or saliva is the gold standard to identify the HCMV infection (Boppana et al., 2011). However, sample collection could be completed within 2–3 weeks after birth to distinguish between congenital and postpartum acquired HCMV infections (Fowler and Boppana, 2018; Pellegrinelli et al., 2020). The causes of postpartum infections are multifaceted, such as exposure to maternal reproductive tract secretions, maternal breast intake, and HCMV serological blood transfusion during delivery (Ross et al., 2011). Postpartum infection is not associated with sensorineural deafness. HCMV infection may be asymptomatic. However, there may be sepsis, including hepatomegaly, splenomegaly, thrombocytopenia, and pneumonia (Park et al., 2013; Chisholm et al., 2014). Based on the dry blood spot test (DBS), the basic PCR is used to detect HCMV DNA. This method includes DNA extraction from the DBS and viral DNA amplification (de Vries et al., 2012). However, studies have shown that DBS PCR has low specificity and sensitivity for neonatal HCMV screening (Boppana et al., 2010; Ross et al., 2017). Therefore, DBS PCR cannot be used as an early diagnostic method to identify most HCMV-infected newborns (Fowler and Boppana, 2018).

Traditionally, isolation of viruses from urine or saliva for tissue culture is the standard method for diagnosing congenital HCMV infection. However, this technology is labor and resource-intensive, and it is time-consuming. Thus, it is not suitable for wide population screening. In comparison, PCR technology is a low cost, fast turnover time, does not require maintenance of tissue culture facilities, and is not easily affected by the storage and transport conditions of samples; therefore, it is suitable for a wide range of neonatal screening (Boppana et al., 2010, 2011; Pinninti et al., 2015). Compared with the urine, the newborn's saliva specimen is easier to collect and is not easily contaminated. The saliva specimens have been proven to be as reliable as urine specimens in diagnosing HCMV. Therefore, the saliva PCR method should be considered for screening (Kadambari et al., 2011; Swanson and Schleiss, 2013; Sahiner et al., 2015; Fowler and Boppana, 2018).

TREATMENT

To date, the most widely used anti-HCMV medication is ganciclovir and/or valganciclovir, which inhibits HCMV replication by disrupting viral DNA synthesis (Lim and Lyall, 2017). Valganciclovir is reserved for congenitally-infected neonates with symptomatic diseases at birth, such as microcephaly, intracranial calcifications, abnormal cerebrospinal fluid index, chorioretinitis, or SNHL. Due to insufficient research evidence, antiviral therapy is generally not recommended for infants with mild birth symptoms under 32 weeks of gestational age or over 30 days of age (Chiopris et al., 2020; Nicloux et al., 2020).

Ganciclovir is the first drug specifically used to treat HCMV infection. It has been proved to be safe, tolerable, and effective for severe organ diseases. Valganciclovir is a prodrug of ganciclovir, and oral valganciclovir is as effective as intravenous ganciclovir with fewer short-term adverse side effects (Xu and Yuan, 2018). Ganciclovir can lead to neutropenia and other toxicities. Kimberlin et al. (2015) also found that ganciclovir has a risk of coronary heart disease and carcinogenesis through animal experiments. Although not yet found in human infections, it is important to communicate this information to parents of infected infants using ganciclovir and/or valganciclovir, and it must be emphasized that ganciclovir does not reverse established CNS injury (Swanson and Schleiss, 2013). Drug toxicity should be closely monitored during long-term antiviral treatment. Also, experiments showed that oral valganciclovir could reduce the risk and have a moderate beneficial effect on hearing for 6 months, and oral valganciclovir can avoid long-term intravenous valganciclovir. It is now recommended to use intravenous ganciclovir only for infants who cannot feed. Once fed, oral valganciclovir is suggested (Kimberlin et al., 2015). The short-term side effects of valganciclovir are neutropenia, thrombocytopenia, anemia, and hepatotoxicity. The occurrence of these undesirable effects may require the temporary or permanent interruption of treatment; they can occur at any time during treatment, requiring regular monitoring of the blood count and liver function throughout its duration. Long-term effects are unknown, but gonadotropins and carcinogenic risks have been observed in animals (Angueyra et al., 2020; Nicloux et al., 2020).

Besides, Ohyama et al. (2019) found that VGCV treatment duration was not associated with differential treatment effects. In a study of Kimberlin et al. (2015), a randomized control trial comparing a 6-week and 6-month regimen of oral VGCV, resulted in similar findings. Although hearing function at 12- and 24-month follow-up was significantly improved or maintained

in the 6-month group, no such difference was evident between the groups at the 6-month time point. Ohyama et al. (2019) found that hearing dysfunction is somewhat reversible in cases of moderate or severe impairment, who may benefit from antiviral therapy. However, this plasticity may already be lost in the "most severe" cases, leaving little hope for improvement. These preliminary findings underscore the need for continued close hearing monitoring of these children (McCrary et al., 2019; Ohyama et al., 2019). Despite antiviral therapy's effectiveness, it is believed that greater efforts should be made to treat HCMV infection, especially if further research is needed to overcome antiviral drugs' toxicity.

PREVENTION

Human cytomegalovirus passes through the placenta after invading the mother, leading to a congenital infection in the fetus. Although natural maternal immunity can reduce the likelihood of congenital fetal infection, it does not completely prevent the occurrence of disease (Nelson et al., 2017).

The placenta of rhesus monkeys was particularly similar in anatomy, immunology, and physiology to humans, which can be used to establish a model of HCMV infection. During the second trimester of pregnancy, by intravenous vaccination with immunological activity, HCMV can cross the placenta and cause congenital HCMV infection, which is then treated by injecting HCMV neutralizing hyperimmune globulin (HIG). The results indicate that effective neutralizing antibodies prevent primary maternal HCMV infection from spreading through the fetus's placenta (Nelson et al., 2017). Researchers have demonstrated that HCMV-HIG could inhibit the spread of HCMV and restore placental health in primary maternal HCMV infection *in vivo* and *in vitro* (Schleiss, 2006; Maidji et al., 2010; McVoy et al., 2018; Penka et al., 2018). However, data from a recent HCMV-HIG randomized trial did not show a significant reduction in fetal infection (Revello et al., 2014). A large multicenter randomized trial of HCMV HIG is currently being conducted in the United States. The results further clarify the role of HIG in preventing congenital HCMV infection (Fowler and Boppana, 2018).

In the absence of effective immunization measures, maternal infection prevention mainly depends on behavioral measures (Manicklal et al., 2013). Infants and young children are more likely to excrete HCMV in saliva and urine than older children and adults. Most pregnant women receive HCMV from children under 3 years of age at home or care for children, while working in nurseries (Adler, 1989; Adler et al., 2004). Educational and behavioral changes during pregnancy can prevent a mother's HCMV infection. However, prevention opportunities are missed because most women have not heard of HCMV or how to prevent it. Therefore, women should be informed about reducing HCMV infection risk during pregnancy (Cannon et al., 2012). Both Adler and Revello's studies have shown that education can effectively prevent vertical transmission and reduce congenital HCMV (Adler, 2015; Revello et al., 2015). Epidemiology showed that HCMV could be transmitted by direct contact with infectious fluids such as urine, saliva, and semen (Cannon et al., 2011). Although sexual contact can spread HCMV, for women of childbearing age, the exposure to urine and saliva of young children is the major transmission route. Avoiding kissing young children and avoiding sharing food and drinks, changing diapers, and washing hands as soon as possible after wiping mouths and nose are all effective behavioral measures (Marsico and Kimberlin, 2017). For seronegative women, effective behavioral prevention can reduce the spread of HCMV. Restricting exposure to HCMV is also beneficial for seropositive women.

Effective preventive measures are significant because of the limited treatment role in altering infants' clinical performance (Nassetta et al., 2009). The development of HCMV vaccines is the most promising strategy for preventing congenital HCMV infection. Studies have shown that effective vaccines are highly cost-effective. It can prevent long-term neurological sequelae and other disabilities by preventing congenital HCMV infection (Dempsey et al., 2012; Swanson and Schleiss, 2013). However, the complex immune avoidance mechanism of HCMV, the relative complexity of its genome, the glycoproteins associated with cell convergence, and the lack of definite HCMV immunogen hinder the development of the HCMV vaccine (Wu et al., 2015). Currently, various programs for developing HCMV vaccines are being developed. Several vaccines can stimulate humoral immunity, while others can enhance natural immunity by stimulating cellular immunity (Griffiths et al., 2013). Although the American Medical Research Institute has given the development of HCMV vaccines top priority, various candidate vaccines have been developed and evaluated in the first phase of clinical trials, and few phase II trials have been successfully conducted. Notably, two vaccines have shown promising results in trials. The gB/MF59 vaccine shows 50% vaccine efficacy in healthy postpartum women. In transplant patients, both gB/MF59 and DNA vaccine TransVax limit viremia periods (Rieder and Steininger, 2014). This study gives researchers confidence in the HCMV vaccine and believes that the HCMV vaccine will be available in the market soon.

CONCLUSION

As the major cause of congenital malformation, HCMV infection can lead to central nervous system sequelae, such as SNHL. These findings in this review suggest that the mechanism of HCMV-related SNHL is still relatively indistinct. At present, more researches are focusing on the mechanism that HCMV invades into the SGN of the inner ear to cause a series of immune responses, which leads to apoptosis of SGN cells and destruction of the inner ear structure. In response to HCMV infection, ganciclovir and/or valganciclovir are still efficient for antiviral treatment. Although progress has been made in the treatment, the detection methods during pregnancy still lack clinically.

The diagnosis of primary HCMV infection during pregnancy is based primarily on serological testing. The mother's primary infection can be determined by detecting the IgG's affinity and the seroconversion from negative to

positive. HCMV can be delivered to the fetus by passing through the placenta. Amniocentesis for viral culture and PCR could be used to determine fetal infection of HCMV. The diagnosis of neonatal HCMV is difficult because most infected individuals are asymptomatic. Isolation of HCMV from urine or saliva is the gold standard for the diagnosis of HCMV infection. Currently, symptomatic SNHL is primarily screened by neonatal hearing screening. Therefore, HCMV prenatal and neonatal screening is important and not comprehensive. Unfortunately, vaccines against HCMV have not been developed to prevent HCMV infection. However, studies have shown that HIG injection can prevent HCMV from infecting the fetus through the placenta. However, the most effective strategy to prevent HCMV vertical transmission is health counseling for women of childbearing age, which is also conducive to screening seronegative women. However, once HCMV infection is detected in pregnant women, there is no effective way to reduce the risk of transmission to the fetus, nor the effective treatment (Mack et al., 2017). Therefore, we advocate preventing HCMV infection by behavioral prevention, such as screening and intervention during pregnancy, not sharing food or drink with children, not kissing children's mouths, and frequent hand washing. A study has shown that behavioral control can reduce maternal HCMV infection and protect newborns from viral infection (Fowler and Boppana, 2018). It is believed that health education for women of childbearing age, hearing screening for newborns after birth, and subsequent HCMV PCR diagnosis of symptomatic infants and symptomatic infants with antiviral treatment can reduce the incidence of SNHL after HCMV infection (Mack et al., 2017). Because HCMV can damage SGN and causes SNHL, further studies are needed to explore the mechanism of SNHL caused by HCMV and provide scientific evidence for follow-up treatment and seek the gospel for most children with HCMV-related SNHL.

AUTHOR CONTRIBUTIONS

WX, HY, CW, CL, and WW collected all the data. WG and YZ designed the figure. ZL wrote the review. All authors read and approved the final manuscript.

FUNDING

This work was funded by grants from the National Natural Science Foundation of China (81471048 and 81541093), the Natural Science Foundation of Shandong Province (ZR2019MC059 and ZR2015HL075), and Shandong Province Innovation and Entrepreneurship Training Program for College Students (S202010438069 and S201910438020X).

REFERENCES

Adler, S. P. (1989). Cytomegalovirus and child day care. Evidence for an increased infection rate among day-care workers. *N. Engl. J. Med.* 321, 1290–1296. doi: 10.1056/NEJM198911093211903

Adler, S. P. (2015). Prevention of maternal-Fetal transmission of cytomegalovirus. *EBioMedicine* 2, 1027–1028. doi: 10.1016/j.ebiom.2015.08.004

Adler, S. P., Finney, J. W., Manganello, A. M., and Best, A. M. (2004). Prevention of child-to-mother transmission of cytomegalovirus among pregnant women. *J. Pediatr.* 145, 485–491. doi: 10.1016/j.jpeds.2004.05.041

Almishaal, A. A., Mathur, P. D., Hillas, E., Chen, L., Zhang, A., Yang, J., et al. (2017). Natural killer cells attenuate cytomegalovirus-induced hearing loss in mice. *PLoS Pathog.* 13:e1006599. doi: 10.1371/journal.ppat.1006599

Angueyra, C., Abou Hatab, H., and Pathak, A. (2020). Congenital cytomegalovirus and Zika infections. *Indian J. Pediatr.* 87, 840–845. doi: 10.1007/s12098-020-03260-9

Bailey, E. M., and Green, S. H. (2014). Postnatal expression of neurotrophic factors accessible to spiral ganglion neurons in the auditory system of adult hearing and deafened rats. *J. Neurosci.* 34, 13110–13126. doi: 10.1523/JNEUROSCI.1014-14.2014

Bartlett, A. W., McMullan, B., Rawlinson, W. D., and Palasanthiran, P. (2017). Hearing and neurodevelopmental outcomes for children with asymptomatic congenital cytomegalovirus infection: a systematic review. *Rev. Med. Virol.* doi: 10.1002/rmv.1938 [Ebup ahead of print]

Bodeus, M., Hubinont, C., Bernard, P., Bouckaert, A., Thomas, K., and Goubau, P. (1999). Prenatal diagnosis of human cytomegalovirus by culture and polymerase chain reaction: 98 pregnancies leading to congenital infection. *Prenat. Diagn.* 19, 314–317. doi: 10.1002/(SICI)1097-0223(199904)19:4<314::AID-PD542>3.0.CO;2-H

Boppana, S. B., Ross, S. A., Novak, Z., Shimamura, M., Tolan, R. W. Jr., Palmer, A. L., et al. (2010). Dried blood spot real-time polymerase chain reaction assays to screen newborns for congenital cytomegalovirus infection. *JAMA* 303, 1375–1382. doi: 10.1001/jama.2010.423

Boppana, S. B., Ross, S. A., Shimamura, M., Palmer, A. L., Ahmed, A., Michaels, M. G., et al. (2011). Saliva polymerase-chain-reaction assay for cytomegalovirus screening in newborns. *N. Engl. J. Med.* 364, 2111–2118. doi: 10.1056/NEJMoa1006561

Bradford, R. D., Yoo, Y. G., Golemac, M., Pugel, E. P., Jonjic, S., and Britt, W. J. (2015). Murine CMV-induced hearing loss is associated with inner ear inflammation and loss of spiral ganglia neurons. *PLoS Pathog.* 11:e1004774. doi: 10.1371/journal.ppat.1004774

Cannon, M. J., Hyde, T. B., and Schmid, D. S. (2011). Review of cytomegalovirus shedding in bodily fluids and relevance to congenital cytomegalovirus infection. *Rev. Med. Virol.* 21, 240–255. doi: 10.1002/rmv.695

Cannon, M. J., Westbrook, K., Levis, D., Schleiss, M. R., Thackeray, R., and Pass, R. F. (2012). Awareness of and behaviors related to child-to-mother transmission of cytomegalovirus. *Prev. Med.* 54, 351–357. doi: 10.1016/j.ypmed.2012.03.009

Carraro, M., Almishaal, A., Hillas, E., Firpo, M., Park, A., and Harrison, R. V. (2017). Cytomegalovirus (CMV) infection causes degeneration of Cochlear vasculature and hearing loss in a mouse model. *J. Assoc. Res. Otolaryngol.* 18, 263–273. doi: 10.1007/s10162-016-0606-4

Chiopris, G., Veronese, P., Cusenza, F., Procaccianti, M., Perrone, S., Dacco, V., et al. (2020). Congenital cytomegalovirus infection: update on diagnosis and treatment. *Microorganisms* 8:1516. doi: 10.3390/microorganisms8101516

Chisholm, K. M., Aziz, N., McDowell, M., Guo, F. P., Srinivas, N., Benitz, W. E., et al. (2014). Evaluation of serial urine viral cultures for the diagnosis of cytomegalovirus infection in neonates and infants. *Pediatr. Dev. Pathol.* 17, 176–180. doi: 10.2350/14-01-1432-OA.1

Cohen-Salmon, M., Regnault, B., Cayet, N., Caille, D., Demuth, K., Hardelin, J. P., et al. (2007). Connexin30 deficiency causes instrastrial fluid-blood barrier disruption within the cochlear stria vascularis. *Proc. Natl. Acad. Sci. U. S. A.* 104, 6229–6234. doi: 10.1073/pnas.0605108104

Davis, N. L., King, C. C., and Kourtis, A. P. (2017). Cytomegalovirus infection in pregnancy. *Birth Defects Res.* 109, 336–346. doi: 10.1002/bdra.23601

Dempsey, A. F., Pangborn, H. M., and Prosser, L. A. (2012). Cost-effectiveness of routine vaccination of adolescent females against cytomegalovirus. *Vaccine* 30, 4060–4066. doi: 10.1016/j.vaccine.2012.04.011

de Vries, J. J., Barbi, M., Binda, S., and Claas, E. C. (2012). Extraction of DNA from dried blood in the diagnosis of congenital CMV infection. *Methods Mol. Biol.* 903, 169–175. doi: 10.1007/978-1-61779-937-2_10

Dhondt, C., Maes, L., Oostra, A., and Dhooge, I. (2019). Episodic vestibular symptoms in children with a congenital cytomegalovirus infection: a case series. *Otol. Neurotol.* 40, e636–e642. doi: 10.1097/MAO.0000000000002244

Dobbie, A. M. (2017). Evaluation and management of cytomegalovirus-associated congenital hearing loss. *Curr. Opin. Otolaryngol. Head Neck Surg.* 25, 390–395. doi: 10.1097/MOO.0000000000000401

Enders, G., Bader, U., Lindemann, L., Schalasta, G., and Daiminger, A. (2001). Prenatal diagnosis of congenital cytomegalovirus infection in 189 pregnancies with known outcome. *Prenat. Diagn.* 21, 362–377. doi: 10.1002/pd.59

Firbas, W., Gruber, H., and Wicke, W. (1981). The blood vessels of the limbus spiralis. *Arch. Otorhinolaryngol.* 232, 131–137. doi: 10.1007/BF00505032

Fowler, K. B., and Boppana, S. B. (2018). Congenital cytomegalovirus infection. *Semin. Perinatol.* 42, 149–154. doi: 10.1053/j.semperi.2018.02.002

Gabrielli, L., Bonasoni, M. P., Santini, D., Piccirilli, G., Chiereghin, A., Guerra, B., et al. (2013). Human fetal inner ear involvement in congenital cytomegalovirus infection. *Acta Neuropathol. Commun.* 1:63. doi: 10.1186/2051-5960-1-63

Gouarin, S., Palmer, P., Cointe, D., Rogez, S., Vabret, A., Rozenberg, F., et al. (2001). Congenital HCMV infection: a collaborative and comparative study of virus detection in amniotic fluid by culture and by PCR. *J. Clin. Virol.* 21, 47–55. doi: 10.1016/S1386-6532(00)00184-0

Griffiths, P., Plotkin, S., Mocarski, E., Pass, R., Schleiss, M., Krause, P., et al. (2013). Desirability and feasibility of a vaccine against cytomegalovirus. *Vaccine* 31(Suppl. 2), B197–B203. doi: 10.1016/j.vaccine.2012.10.074

Hammer, Q., Ruckert, T., and Romagnani, C. (2018). Natural killer cell specificity for viral infections. *Nat. Immunol.* 19:800–808. doi: 10.1038/s41590-018-0163-6

Juckstock, J., Rothenburger, M., Friese, K., and Traunmuller, F. (2015). Passive immunization against congenital cytomegalovirus infection: current state of knowledge. *Pharmacology* 95, 209–217. doi: 10.1159/000381626

Juhn, S. K., and Rybak, L. P. (1981). Labyrinthine barriers and cochlear homeostasis. *Acta Otolaryngol.* 91, 529–534. doi: 10.3109/00016488109138538

Kadambari, S., Williams, E. J., Luck, S., Griffiths, P. D., and Sharland, M. (2011). Evidence based management guidelines for the detection and treatment of congenital CMV. *Early Hum. Dev.* 87, 723–728. doi: 10.1016/j.earlhumdev.2011.08.021

Kimberlin, D. W., Jester, P. M., Sanchez, P. J., Ahmed, A., Arav-Boger, R., Michaels, M. G., et al. (2015). Valganciclovir for symptomatic congenital cytomegalovirus disease. *N. Engl. J. Med.* 372, 933–943. doi: 10.1056/NEJMoa1404599

Kimura, R. S., Nye, C. L., and Southard, R. E. (1990). Normal and pathologic features of the limbus spiralis and its functional significance. *Am. J. Otolaryngol.* 11, 99–111. doi: 10.1016/0196-0709(90)90006-H

Li, X., Shi, X., Qiao, Y., Xu, K., Zeng, L., Wang, C., et al. (2014). Observation of permeability of blood-labyrinth barrier during cytomegalovirus-induced hearing loss. *Int. J. Pediatr. Otorhinolaryngol.* 78, 995–999. doi: 10.1016/j.ijporl.2014.03.013

Li, X., Shi, X., Wang, C., Niu, H., Zeng, L., and Qiao, Y. (2016). Cochlear spiral ganglion neuron apoptosis in neonatal mice with murine cytomegalovirus-induced Sensorineural hearing loss. *J. Am. Acad. Audiol.* 27, 345–353. doi: 10.3766/jaaa.15061

Lim, Y., and Lyall, H. (2017). Congenital cytomegalovirus - who, when, what-with and why to treat? *J. Infect.* 74(Suppl. 1), S89–S94. doi: 10.1016/S0163-4453(17)30197-4

Lipitz, S., Elkan Miller, T., Yinon, Y., Weissbach, T., De-Castro, H., Hoffman, C., et al. (2020). Revisiting short- and long-term outcome after fetal first-trimester primary cytomegalovirus infection in relation to prenatal imaging findings. *Ultrasound Obstet. Gynecol.* 56, 572–578. doi: 10.1002/uog.21946

Lipitz, S., Hoffmann, C., Feldman, B., Tepperberg-Dikawa, M., Schiff, E., and Weisz, B. (2010). Value of prenatal ultrasound and magnetic resonance imaging in assessment of congenital primary cytomegalovirus infection. *Ultrasound Obstet. Gynecol.* 36, 709–717. doi: 10.1002/uog.7657

Mack, I., Burckhardt, M. A., Heininger, U., Prufer, F., Schulzke, S., and Wellmann, S. (2017). Symptomatic congenital cytomegalovirus infection in children of seropositive women. *Front. Pediatr.* 5:134. doi: 10.3389/fped.2017.00134

Maidji, E., Nigro, G., Tabata, T., McDonagh, S., Nozawa, N., Shiboski, S., et al. (2010). Antibody treatment promotes compensation for human cytomegalovirus-induced pathogenesis and a hypoxia-like condition in placentas with congenital infection. *Am. J. Pathol.* 177, 1298–1310. doi: 10.2353/ajpath.2010.091210

Manicklal, S., Emery, V. C., Lazzarotto, T., Boppana, S. B., and Gupta, R. K. (2013). The "silent" global burden of congenital cytomegalovirus. *Clin. Microbiol. Rev.* 26, 86–102. doi: 10.1128/CMR.00062-12

Marsico, C., and Kimberlin, D. W. (2017). Congenital cytomegalovirus infection: advances and challenges in diagnosis, prevention and treatment. *Ital. J. Pediatr.* 43:38. doi: 10.1186/s13052-017-0358-8

McCrary, H., Sheng, X., Greene, T., and Park, A. (2019). Long-term hearing outcomes of children with symptomatic congenital CMV treated with valganciclovir. *Int. J. Pediatr. Otorhinolaryngol.* 118, 124–127. doi: 10.1016/j.ijporl.2018.12.027

McVoy, M. M., Tenorio, E., and Kauvar, L. M. (2018). A native human monoclonal antibody targeting HCMV gB (AD-2 site I). *Int. J. Mol. Sci.* 19:3982. doi: 10.3390/ijms19123982

Mestas, E. (2016). Congenital cytomegalovirus. *Adv. Neonatal Care* 16, 60–65. doi: 10.1097/ANC.0000000000000242

Mittal, R., Aranke, M., Debs, L. H., Nguyen, D., Patel, A. P., Grati, M., et al. (2017). Indispensable role of ion channels and transporters in the auditory system. *J. Cell. Physiol.* 232, 743–758. doi: 10.1002/jcp.25631

Moteki, H., Isaka, Y., Inaba, Y., Motobayashi, M., Nishio, S. Y., Ohira, S., et al. (2018). A rational approach to identifying newborns with hearing loss caused by congenital cytomegalovirus infection by dried blood spot screening. *Acta Otolaryngol.* 138, 708–712. doi: 10.1080/00016489.2018.1441545

Naing, Z. W., Scott, G. M., Shand, A., Hamilton, S. T., van Zuylen, W. J., Basha, J., et al. (2016). Congenital cytomegalovirus infection in pregnancy: a review of prevalence, clinical features, diagnosis and prevention. *Aust. N. Z. J. Obstet. Gynaecol.* 56, 9–18. doi: 10.1111/ajo.12408

Nassetta, L., Kimberlin, D., and Whitley, R. (2009). Treatment of congenital cytomegalovirus infection: implications for future therapeutic strategies. *J. Antimicrob. Chemother.* 63, 862–867. doi: 10.1093/jac/dkp083

Navti, O. B., Al-Belushi, M., and Konje, J. C.Frcog (2020). Cytomegalovirus infection in pregnancy - an update. *Eur. J. Obstet. Gynecol. Reprod. Biol.* 258, 216–222. doi: 10.1016/j.ejogrb.2020.12.006

Nelson, C. S., Cruz, D. V., Tran, D., Bialas, K. M., Stamper, L., Wu, H., et al. (2017). Preexisting antibodies can protect against congenital cytomegalovirus infection in monkeys. *JCI Insight* 2:e94002. doi: 10.1172/jci.insight.94002

Nicloux, M., Peterman, L., Parodi, M., and Magny, J. F. (2020). Outcome and management of newborns with congenital cytomegalovirus infection. *Arch. Pediatr.* 27, 160–165. doi: 10.1016/j.arcped.2020.01.006

Ohyama, S., Morioka, I., Fukushima, S., Yamana, K., Nishida, K., Iwatani, S., et al. (2019). Efficacy of valganciclovir treatment depends on the severity of hearing dysfunction in symptomatic infants with congenital cytomegalovirus infection. *Int. J. Mol. Sci.* 20:1388. doi: 10.3390/ijms20061388

Park, A. H., Mann, D., Error, M. E., Miller, M., Firpo, M. A., Wang, Y., et al. (2013). Comparative analysis of detection methods for congenital cytomegalovirus infection in a Guinea pig model. *JAMA Otolaryngol. Head Neck Surg.* 139, 82–86. doi: 10.1001/jamaoto.2013.1090

Pellegrinelli, L., Alberti, L., Pariani, E., Barbi, M., and Binda, S. (2020). Diagnosing congenital cytomegalovirus infection: don't get rid of dried blood spots. *BMC Infect. Dis.* 20:217. doi: 10.1186/s12879-020-4941-z

Penka, L., Kagan, K. O., Goelz, R., and Hamprecht, K. (2018). Comparison of quantitative real-time PCR and short-term (18-hour) microculture in diagnosis of fetal cytomegalovirus infection: impact of hyperimmunoglobulin treatment. *Prenat. Diagn.* 38, 936–942. doi: 10.1002/pd.5338

Pinninti, S. G., Ross, S. A., Shimamura, M., Novak, Z., Palmer, A. L., Ahmed, A., et al. (2015). Comparison of saliva PCR assay versus rapid culture for detection of congenital cytomegalovirus infection. *Pediatr. Infect. Dis. J.* 34, 536–537. doi: 10.1097/INF.0000000000000609

Prince, H. E., and Lape-Nixon, M. (2014). Role of cytomegalovirus (CMV) IgG avidity testing in diagnosing primary CMV infection during pregnancy. *Clin. Vaccine Immunol.* 21, 1377–1384. doi: 10.1128/CVI.00487-14

Revello, M. G., and Gerna, G. (2002). Diagnosis and management of human cytomegalovirus infection in the mother, fetus, and newborn infant. *Clin. Microbiol. Rev.* 15, 680–715. doi: 10.1128/CMR.15.4.680-715.2002

Revello, M. G., Lazzarotto, T., Guerra, B., Spinillo, A., Ferrazzi, E., Kustermann, A., et al. (2014). A randomized trial of hyperimmune globulin to prevent congenital cytomegalovirus. *N. Engl. J. Med.* 370, 1316–1326. doi: 10.1056/NEJMoa1310214

Revello, M. G., Lilleri, D., Zavattoni, M., Furione, M., Middeldorp, J., and Gerna, G. (2003). Prenatal diagnosis of congenital human cytomegalovirus infection in amniotic fluid by nucleic acid sequence-based amplification assay. *J. Clin. Microbiol.* 41, 1772–1774. doi: 10.1128/JCM.41.4.1772-1774.2003

Revello, M. G., Tibaldi, C., Masuelli, G., Frisina, V., Sacchi, A., Furione, M., et al. (2015). Prevention of primary cytomegalovirus infection in pregnancy. *EBioMedicine* 2, 1205–1210. doi: 10.1016/j.ebiom.2015.08.003

Rieder, F., and Steininger, C. (2014). Cytomegalovirus vaccine: phase II clinical trial results. *Clin. Microbiol. Infect.* 20(Suppl. 5), 95–102. doi: 10.1111/1469-0691.12449

Ross, S. A., Ahmed, A., Palmer, A. L., Michaels, M. G., Sanchez, P. J., Stewart, A., et al. (2017). Newborn dried blood spot polymerase chain reaction to identify infants with congenital cytomegalovirus-associated sensorineural hearing loss. *J. Pediatr.* 184, 57.e1–61.e1. doi: 10.1016/j.jpeds.2017.01.047

Ross, S. A., Novak, Z., Pati, S., and Boppana, S. B. (2011). Overview of the diagnosis of cytomegalovirus infection. *Infect. Disord. Drug Targets* 11, 466–474. doi: 10.2174/187152611797636703

Sahiner, F., Cekmez, F., Cetinkaya, M., Kaya, G., Kalayci, T., Gunes, O., et al. (2015). Congenital cytomegalovirus infections and glycoprotein B genotypes in live-born infants: a prevalence study in Turkey. *Infect. Dis.* 47, 465–471. doi: 10.3109/23744235.2015.1018316

Schleiss, M. R. (2006). The role of the placenta in the pathogenesis of congenital cytomegalovirus infection: is the benefit of cytomegalovirus immune globulin for the newborn mediated through improved placental health and function? *Clin. Infect. Dis.* 43, 1001–1003. doi: 10.1086/507642

Schmutzhard, J., Glueckert, R., Pritz, C., Blumer, M. J., Bitsche, M., Lackner, P., et al. (2013). Sepsis otopathy: experimental sepsis leads to significant hearing impairment due to apoptosis and glutamate excitotoxicity in murine cochlea. *Dis. Model. Mech.* 6, 745–754. doi: 10.1242/dmm.011205

Shi, X., Qiu, S., Zhuang, W., Yuan, N., Wang, C., Zhang, S., et al. (2017). NLRP3-inflammasomes are triggered by age-related hearing loss in the inner ear of mice. *Am. J. Transl. Res.* 9, 5611–5618.

Swanson, E. C., and Schleiss, M. R. (2013). Congenital cytomegalovirus infection: new prospects for prevention and therapy. *Pediatr. Clin. N. Am.* 60, 335–349. doi: 10.1016/j.pcl.2012.12.008

Teissier, N., Bernard, S., Quesnel, S., and Van Den Abbeele, T. (2016). Audiovestibular consequences of congenital cytomegalovirus infection. *Eur. Ann. Otorhinolaryngol. Head Neck Dis.* 133, 413–418. doi: 10.1016/j.anorl.2016.03.004

Tian, G., Zhang, S., and Yang, J. (2018). Coexistence of IL-6-572C/G and ICAM-1 K469E polymorphisms among patients with sudden sensorineural hearing loss. *Tohoku J. Exp. Med.* 245, 7–12. doi: 10.1620/tjem.245.7

Wentland, C. J., Ronner, E. A., Basonbul, R. A., Pinnapureddy, S., Mankarious, L., Keamy, D., et al. (2018). Utilization of diagnostic testing for pediatric sensorineural hearing loss. *Int. J. Pediatr. Otorhinolaryngol.* 111, 26–31. doi: 10.1016/j.ijporl.2018.05.024

Wu, J., Han, W., Chen, X., Guo, W., Liu, K., Wang, R., et al. (2017). Matrix metalloproteinase-2 and -9 contribute to functional integrity and noiseinduced damage to the blood-labyrinth-barrier. *Mol. Med. Rep.* 16, 1731–1738. doi: 10.3892/mmr.2017.6784

Wu, S. J., Villarreal, D. O., Shedlock, D. J., and Weiner, D. B. (2015). Synthetic DNA approach to cytomegalovirus vaccine/immune therapy. *Adv. Exp. Med. Biol.* 848, 131–148. doi: 10.1007/978-1-4939-2432-5_7

Xu, W. F., and Yuan, T. M. (2018). A review on the prevention and treatment of congenital cytomegalovirus infection in mothers and infants. *Zhongguo Dang Dai Er Ke Za Zhi* 20, 870–875. doi: 10.7499/j.issn.1008-8830.2018.10.018

Yinon, Y., Farine, D., and Yudin, M. H. (2018). No. 240-cytomegalovirus infection in pregnancy. *J. Obstet. Gynaecol. Can.* 40, e134–e141. doi: 10.1016/j.jogc.2017.11.018

Zhao, H. B. (2017). Hypothesis of K^+-recycling defect is not a primary deafness mechanism for Cx26 (GJB2) deficiency. *Front. Mol. Neurosci.* 10:162. doi: 10.3389/fnmol.2017.00162

Zhuang, W., Wang, C., Shi, X., Qiu, S., Zhang, S., Xu, B., et al. (2018). MCMV triggers ROS/NLRP3-associated inflammasome activation in the inner ear of mice and cultured spiral ganglion neurons, contributing to sensorineural hearing loss. *Int. J. Mol. Med.* 41, 3448–3456. doi: 10.3892/ijmm.2018.3539

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Microbiology

ORIGINAL RESEARCH
published: 03 May 2021
doi: 10.3389/fmicb.2021.675112

The Second-Generation XPO1 Inhibitor Eltanexor Inhibits Human Cytomegalovirus (HCMV) Replication and Promotes Type I Interferon Response

*Yueyan Liao[1], Xiangyu Ke[1†], Tianyi Deng[1] and Qingsong Qin[1,2,3]**

[1] Laboratory of Human Virology and Oncology, Shantou University Medical College, Shantou, China, [2] Guangdong Provincial Key Laboratory of Infectious Diseases and Molecular Immunopathology, Shantou, China, [3] Guangdong Provincial Key Laboratory for Diagnosis and Treatment of Breast Cancer, Shantou, China

Human cytomegalovirus (HCMV) is a ubiquitous opportunistic pathogen and can be life-threatening for immunocompromised individuals. There is currently no available vaccine for the prevention of HCMV- associated diseases and most of the available antiviral drugs that target viral DNA synthesis become ineffective in treating HCMV mutants that arise after long-term use in immunocompromised patients. Here, we examined the effects of Eltanexor, a second-generation selective inhibitor of nuclear export (SINE), on HCMV replication. Eltanexor effectively inhibits HCMV replication in human foreskin fibroblasts in a dose-dependent manner. Eltanexor does not significantly inhibit viral entry and nuclear import of viral genomic DNA, but rather suppress the transcript and protein levels of viral immediate-early (IE), early (E) and late (L) genes, and abolishes the production of infectious virions. We further found Eltanexor treatment promotes proteasome-mediated degradation of XPO1, which contributes to the nuclear retention of interferon regulatory factor 3 (IRF-3), resulting in increased expression of type I interferon as well as interferon stimulating genes ISG15 and ISG54. This study reveals a novel antiviral mechanism of Eltanexor which suggests it has potential to inhibit a broad spectrum of viral pathogens.

Keywords: cytomegalovirus, type I interferon, replication, Eltanexor, XPO1 inhibitor

OPEN ACCESS

Edited by:
Julien Andreani,
IHU Mediterranee Infection, France

Reviewed by:
Michael Nevels,
University of St. Andrews,
United Kingdom
Beatrice Mercorelli,
University of Padua, Italy

***Correspondence:**
Qingsong Qin
qsqin@stu.edu.cn

†Present address:
Xiangyu Ke,
Centre of Clinical Epidemiology and Methodology, Guangdong Second Provincial General Hospital, Guangzhou, China

Specialty section:
This article was submitted to Virology, a section of the journal Frontiers in Microbiology

Received: *02 March 2021*
Accepted: *12 April 2021*
Published: *03 May 2021*

Citation:
Liao Y, Ke X, Deng T and Qin Q (2021) The Second-Generation XPO1 Inhibitor Eltanexor Inhibits Human Cytomegalovirus (HCMV) Replication and Promotes Type I Interferon Response. Front. Microbiol. 12:675112. doi: 10.3389/fmicb.2021.675112

INTRODUCTION

Human cytomegalovirus (HCMV) is a β-herpesvirus, which infects up to 60–90% of the world's population. HCMV infection is generally asymptomatic but can cause life-threatening complications in immunocompromised individuals, such as transplant recipients and patients with AIDS. It is also a leading cause of congenital infections (Cannon, 2019). Currently there is no vaccine for HCMV, and effects of antiviral therapy are limited due to the emergence of resistant

viral mutants during a long-term treatment (Poole and James, 2018). Most antiviral agents are nucleoside analogs that target viral DNA synthesis. Unfortunately, antiviral therapy is hindered by the emergence of mutations in viral DNA polymerase UL54 and viral kinase UL97 which confer resistance to common anti-HCMV drugs such as ganciclovir and maribavir (Razonable, 2018). The newly FDA-approved anti-HCMV drug, letermovir, which targets the viral DNA terminase-complex, has been in use for a little over a year; however, resistant viral variants have already emerged (Douglas et al., 2020). Thus, there remains an urgent need to develop new antiviral therapies for HCMV.

Viruses are obligate intracellular agents and rely on cellular trafficking systems. Targeting the nuclear export machinery is a potential alternative strategy for antiviral therapies (Mathew and Ghildyal, 2017; Kosyna and Depping, 2018). Exportins are Ran-GTPase-dependent nuclear transport factors that belong to the karyopherin-β family and mediate the nuclear export of a plethora of proteins and RNAs (Matsuura, 2016). Exportin-1 (XPO1), also known as chromosome region maintenance 1 (CRM1), is a major receptor responsible for exporting proteins from the nucleus to the cytoplasm. XPO1 mediates nuclear export of more than 220 proteins including tumor suppressors and growth regulators such as p53, p21, FOXO, PI3K/AKT, Wnt/β-catenin, AP-1 and NF-κB, and is involved in the regulation of cell cycle progression and apoptosis (Fornerod et al., 1997; Xu et al., 2012). XPO1 binds to a diverse array of cargos in the presence of Ran-GTP via a nuclear export sequence (NES) composed of a cluster of leucine (L)-rich or hydrophobic amino acids (Dong et al., 2009; Monecke et al., 2009). Leptomycin B (LMB) is a prototypical inhibitor of XPO1 and covalently binds to Cys528 located in the NES binding groove of XPO1 (Kudo et al., 1999; Petosa et al., 2004; Monecke et al., 2009; Sun et al., 2013; Turner et al., 2014), and has been studied as a potent agent against various types of cancer as well as an antiviral agent for many years (Mathew and Ghildyal, 2017). However, LMB is unsuitable for therapeutic development, because it irreversibly shuts down nuclear export resulting in high cytotoxicity *in vivo* (Sun et al., 2013; London et al., 2014). This has led to the development of synthetic analogs of LMB (known as the second-generation selective inhibitors of nuclear export [SINEs]), such as KPT8602 (Eltanexor), KPT330 (Selinexor), KPT335 (Verdinexor), KPT185, that have substantially improved *in vivo* tolerance and are reversible (Ranganathan et al., 2012; Azmi et al., 2013; Gutierrez et al., 2013; Zhang et al., 2013; Zheng et al., 2014), and have been extensively tested in phase I/II clinical trials for solid tumors, and hematologic malignancies (Cornell et al., 2016; Hing et al., 2016).

SINEs have also been extensively studied for antiviral therapies as many viruses exploit or modulate XPO1-mediated nuclear export at various stages of their lifecycles (Gruffaz et al., 2019). It has been reported that SINEs inhibit the replication of numerous viruses including influenza virus (Perwitasari et al., 2014), HIV (Boons et al., 2015), Epstein-Barr virus, human cytomegalovirus, Kaposi's sarcoma virus, adenoviruses, BK virus, John Cunningham virus, and human papillomavirus (Widman et al., 2018). However, the antiviral mechanism of SINEs remains to be further studied. A previous study showed that LMB inhibits HCMV replication by blocking the nucleocytoplasmic trafficking of HCMV structural proteins (pp65 and UL94) (Sanchez et al., 2007; Liu et al., 2012). In this study, we examined the effects of Eltanexor (KPT-8602), a newly developed selective inhibitors of nuclear export which showed improved efficacy and *in vivo* tolerability in clinical trials of hematological malignancies (Hing et al., 2016), on HCMV replication. Our results indicate that Eltanexor significantly inhibits HCMV transcript and protein levels during viral lytic infection in fibroblasts. Additionally, Eltanexor targets XPO1 for proteasome-mediated degradation and results in enhanced expression of IFN-β These findings reveal a novel antiviral mechanism of Eltanexor.

MATERIALS AND METHODS

Cells and Viruses

Human foreskin fibroblasts (HFFs) (CRL-4001, ATCC, passages: 10–20) were cultured in Dulbecco's modified Eagle's medium (DMEM) supplemented with 10% fetal bovine serum, 100 U/ml penicillin, and 100 μg/ml streptomycin in an incubator with 5% CO_2 at 37°C. The HCMV strain used was rescued from the HCMV bacterial artificial chromosome (BAC) cosmid termed TB40/E*wt*-mCherry (a generous gift from Eain Murphy at SUNY Upstate Medical University), which has a mCherry marker gene inserted between US34A and TSR1 and replicates like its wild type parent virus as previously described (Sinzger et al., 2008; O'Connor and Shenk, 2011). Viral lysates were pelleted through a D-sorbitol cushion (20%, weight/volume), centrifuged at 20,000 rpm (Hitachi) (equal to 41,224 × g) for 1 h, at 20°C. Viral pellets were resuspended and aliquoted in DMEM, and further titrated by a $TCID_{50}$ assay. To better explain the multiplicity of infection (MOI) in infection assays, viral titers generated by TCID50 assays were expressed as PFU based the formula (1 PFU = 0.69 * TCID50).

Chemicals and Antibodies

Chemicals include Eltanexor (HY-100423, MCE, China) with a purity of 99.71%, polyinosine-polycytidylic acid [poly(I:C)] (B5551, APExBIO Technology, TX, United States), and the proteasome inhibitor MG-132 (S1748, Beyotime, Shanghai, China). Primary antibodies used include: mouse anti-pp52 (UL44) antibodies (0897, sc-58117, Santa Cruz, Dallas, TX, United States), mouse anti-IE1 monoclonal antibody (1B12) and mouse anti-IE2 monoclonal antibody (3H9) (a generous gift from Thomas Shenk at Princeton University), mouse anti-pp65 antibodies (1-L-11, sc-52401, Santa Cruz), mouse anti-pp28 antibodies (CA004, Virusys, MD, United States), rabbit anti-IRF-3 antibodies (11904, Cell Signaling Technologies [CST], Danvers, MA, United States), rabbit anti β-actin antibody (TA-09, ZSGS-Bio, Beijing, China), rabbit anti-pp71 serum (prepared by Chempeptide, Shanghai, China), and rabbit anti-XPO1 monoclonal antibody (D6V7N, CST). Secondary antibodies include horseradish peroxidase (HRP)-conjugated goat anti-mouse immunoglobulin antibodies (7076, CST), HRP-conjugated goat anti-rabbit immunoglobulin antibodies

(7074, CST), FITC-conjugated goat anti-rabbit immunoglobulin antibody (SA00003, Proteintech, IL, United States), and Cy3-conjugated goat anti-mouse immunoglobulin antibody (SA00001, Proteintech).

Cell Viability Assay

HFFs (1×10^4 cells/well) were seeded in 96-well plates and then treated with various concentrations of Eltanexor as indicated and incubated at 37°C for 72 h. Eltanexor was dissolved in dimethyl sulfoxide (DMSO), and an equal volume of DMSO served as a vehicle control. Cell viability was determined by a 3-(4, 5, dimethyliazol-2-yl)-5-(3carboxymethoxy-phenyl)-2-(4-sulfophenyl)-2H-tetrazolium salt (MTS)-based colorimetric assay (Promega) as described by manufacturer's instructions (Meloni et al., 2015). 50% cytotoxic concentration (CC_{50}) is determined according to non-linear trajectory analysis using GraphPad Prism Software (San Diego, CA).

Eltanexor Add-On and Removal Assays

HFFs (1×10^5 cells/well) were seeded in 12-well plates and infected with HCMV at a MOI = 1. For add-on assays, Eltanexor (0.4 μM) was added at 0, 6, 12, 24, 36, 48, and 60 h post infection (hpi). At 72 hpi, drug-containing medium was replaced with normal drug-free medium. At 96 hpi, culture supernatants were collected for viral titration by a $TCID_{50}$ assay on HFFs. For removal assays, HCMV infected cells were pretreated in medium containing Eltanexor (0.4 μM), and Eltanexor was removed by replacing with normal drug-free medium at 6, 12, 24, 36, 48, 60, and 72 hpi. Culture supernatants were collected at 96 hpi for viral titration by a $TCID_{50}$ assay.

Quantification of Viral Genome, Viral, and Cellular Transcripts

(i) Quantification of viral genomic DNA from HCMV infected cells. HFFs (3×10^5 cells/well, 6 well plate) were infected with HCMV at a MOI = 1, treated with Eltanexor or vehicle DMSO at indicated concentrations for 1.5 h on ice. Cells were washed with phosphate-buffered saline (PBS) buffer (137 mM NaCl, 3 mM KCl, 8 mM Na2HPO4 [pH 7.5]), trypsinized for 3 min, and centrifuged at 4℃ for 3 min at 500 × g. Cells were washed twice with cold PBS and subjected to DNA extraction with DNA extraction kit (D4035, OMEGA, GA, United States). Viral relative genomic copy numbers were quantified by qPCR with UL122 primers and a TB green qPCR Kit (RR420A, TAKARA, Dalian, China). For quantification of viral genomic DNA from the nucleus, cytoplasm/nuclear fractions were isolated with a nuclear and cytoplasmic purification kit (HR0241, BJBALB, Beijing, China) based on manufacturer's instructions. Viral DNA from cytoplasm/nuclear fractions was extracted and quantified as above. (ii) Quantification of transcripts of viral and cellular genes. Total RNA was extracted from infected HFFs at indicated times with Trizol (9109, TAKARA, Dalian, China). Equivalent amounts (0.1 μg) of RNA were reverse transcribed into cDNA with universal pentamers using a reverse transcription kit (RR820A, TAKARA, Dalian, China). Transcripts of viral genes (UL36, UL122, UL123, UL44, UL82, UL83, UL99, and UL75) and cellular genes (IFN-β1, ISG15, ISG54) were further quantified by qPCR. β-actin served as an internal control. All primers used are listed in **Table 1**.

Western Blotting Assay

Cells were lysed in RIPA buffer (50mM Tris (pH 7.4), 150mM NaCl, 1% Triton X-100, 1% sodium deoxycholate, 0.1% SDS) (BL504A, Biosharp, China) supplemented with proteinase inhibitors (4693159001, Roche, Basel, Switzerland). Equal amounts of proteins were separated by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE), transferred and immobilized on polyvinylidene fluoride (PVDF) membranes. The membrane was blocked with 5% skim milk in PBST (PBS plus 0.05% Tween 20) for 1 h, followed by overnight incubation with rabbit or mouse primary antibodies for the proteins of interest. Blots were washed three times with PBST for 15 min each, followed by incubation with HRP-conjugated secondary antibodies for 2 h. Blots were then washed with PBST three times and developed with the ECL enhanced chemiluminescence system (G3308, Gbcbio Technologies Inc., Guangzhou, China). Images were taken and analyzed by Amersham Imager 600 system (GE healthcare, United States).

Indirect Immunofluorescence Assay

HFFs (1×10^5 cell/well, 12 well plate) were seeded on coverslips and infected with HCMV at a MOI = 1, fixed at indicated times with 4% paraformaldehyde for 15 min at room temperature, and then permeabilized with 0.1% Triton X-100 in PBS for 15 mins.

TABLE 1 | Primers used in this study.

Primer	Sequence
UL123(F)	5′ GCCTTCCCTAAGACCACCAAT 3′
UL123(R)	5′ ATTTTCTGGGCATAAGCCATAATC 3′
UL44(F)	5′ TACAACAGCGTGTCGTGCTCCG 3′
UL44(R)	5′ GGCGTGAAAAACATGCGTATCAAC 3′
UL99(F)	5′ GTGTCCCATTCCCGACTCG 3′
UL99(R)	5′ TTCACAACGTCCACCCACC 3′
UL122(F)	5′ TGTTCCGTCACACCAATCGTTCTC 3′
UL122(R)	5′ AGGCGACACCGTACCTGATCC 3′
UL36(F)	5′ GGCACCGTCTGTTCGCAAGG 3′
UL36(R)	5′ CCGATGAGCACGATGAGTTGGTAG 3′
UL75-F	5′ TCTTGACGCCGCACTTGACTTC 3′
UL75-R	5′ CATCTGACATCGACCGCTCTTGAG 3′
UL82-F	5′ GCGAGCCTTGACGACTTGGTAC 3′
UL82-R	5′ GAAGTGGAAGCGGTGCTGATGG 3′
Actin(F)	5′ TCCTCCTGAGCGCAAGTACTC 3′
Actin(R)	5′ CGGACTCGTCATACTCCTGCTT 3′
UL83(F)	5′ AGGTGCAGCACACGTACTTT 3′
UL83(R)	5′ TAGTGGTGCACGTTGATGCT 3′
IFNβ1(F)	5′AAACTCATGAGCAGTCTGCA 3′
IFNβ1(R)	5′AGGAGATCTTCAGTTTCGGAGG 3′
ISG15(F)	5′ ATGGGCTGGGACCTGACG 3′
ISG15(R)	5′ GCCAATCTTCTGGGTGATCTG 3′
ISG54(F)	5′ AACCTACTGGCCTATCTAAAGC 3′
ISG54(R)	5′ CATGCTCTTGCTGGATTAACTC 3′

The coverslips were then washed three times with PBST before being blocked with PBST containing 5% bovine serum albumin for 30 min. The cells were then incubated with primary antibodies (mouse anti-pp65, rabbit anti-IRF3) for 1 h. After washing three times with PBST, cells were incubated for another hour with FITC or Cy3-conjugated secondary antibodies. Slides were washed three times with PBST and mounted with anti-fade reagent with DAPI (4', 6-diamidino-2-phenylindole dihydrochloride) (28718-90-3, Biofroxx, United Kingdom). Images were taken with a DeltaVision Elite microscope (GE healthcare, United States).

Enzyme-Linked Immunosorbent Assay (ELISA)

HFFs (1×10^5 cells/well) were seeded in 12-well plates and infected with HCMV at MOI = 1 or mock infected on ice for 1.5 h. Cells were then treated with Eltanexor (0.4 μM) or DMSO. Similarly, HFFs (1×10^5 cells/well) were treated with Poly (I:C) (20μg/mL) with or without Eltanexor (0.4 μM), and then culture supernatants were collected at indicated time points. Production of IFN-β in supernatants was determined with an ELISA kit (KE1364, ImmunoWay, TX, United States).

Statistical Analysis

Statistical significance was assessed with GraphPad Prism 5.0 software. Student's *t*-tests were used to determine differences between the means of two groups. Multiple-comparison significance was determined by two-way ANOVAs. In all analyses, two-sided *p*-values were used, and $p < 0.05$ was considered statistically significant, $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.

RESULTS

Eltanexor Inhibits HCMV Lytic Replication

Eltanexor (KPT-8602) is a newly developed synthetic second-generation XPO1 inhibitor (**Figure 1A**) and less toxic than analogs, and is currently in phase I/II clinical trials for multiple myeloma (Cornell et al., 2016). Therefore, we aim to examine its effects on HCMV replication. Firstly, we analyzed the toxicity of Eltanexor on HFFs. Eltanexor does not significantly affect cell viability at concentrations less than or equal to 0.8 μM (**Figure 1B**), and 50% cytotoxic concentration (CC_{50}) is determined at 14.06 μM (**Figure 1C**). Therefore, we examined the effect of Eltanexor on HCMV replication at concentrations between 0 and 0.8 μM as indicated. Eltanexor inhibits the production of HCMV progeny virions in a dose-dependent manner, and the half-maximal inhibitory concentration (IC_{50} or EC_{50}) is determined at 0.03762 μM (**Figure 1D**). Selectivity of Index (SI) of Eltanxor is calculated as 374. Western-blotting assays also show Eltanexor treatment inhibits the expression of IE2/86 (encoded by UL122), early protein pp52 (UL44), and late proteins pp71 and pp65 (UL82 and UL83) in a dose-dependent manner (**Figure 1E**). Taken together, these results demonstrate that Eltanexor inhibits HCMV replication in a dose-dependent manner.

Eltanexor Does Not Affect Viral Entry or Nuclear Import of Genomic DNA

HCMV replication is a complex process consisting of multiple steps, including attachment and entry, intracellular trafficking, uncoating, genomic replication, protein synthesis, capsid assembly and DNA encapsidation, translocation, envelopment, and egress (Edward et al., 2013). To determine at which step of HCMV replication was interrupted by Eltanexor treatment, Eltanexor add-on and removal assays were performed. The results of add-on assays indicate the production of progeny virions significantly escapes the suppression of Eltanexor treatment when no drug is added before 24 hpi; on the other hand, the results of removal assays indicate the production of progeny virions is inhibited when Eltanexol is removed from Eltanexor pretreated medium later than 36 hpi (**Figure 2A**). To further examine whether Eltanexor affects viral entry at early stages of HCMV infection, localization of viral tegument protein pp65 delivered as part of the virions was determined by immunofluorescence assay. The nuclear translocation of pp65 is not significantly affected by Eltanexor treatment (**Figure 2B**). Quantification of viral genomic DNA in Eltanexor treated HFFs shows viral attachment is not significantly affected in a dose-dependent manner (**Figure 2C**). Furthermore, quantifications of viral genomic DNA isolated from nuclear fractions and wholes cell lysates show that Eltanexor does not affect nuclear import of viral genomic DNA (**Figure 2D**). From these results, we conclude that Eltanexor does not significantly affect viral entry and nuclear import of viral genomic DNA but can exert inhibitory effects within 24–36 hpi.

Eltanexor Treatment Decrease the Transcript and Protein Levels of Viral Immediate-Early (IE2), Early (E), and Late (L) Genes

Since Eltanexor does not affect HCMV entry and trafficking, we speculate that it might affect viral transcript and protein levels. Hence, we examined the transcript levels of viral genes under Eltanexor treatment (0.4 μM). The transcript levels of IE genes (UL36, UL122, and UL123), and E (UL44) and L genes (UL82, UL83, UL99, and UL75) (Landolfo et al., 2003) were quantified by RT-qPCR at 6, 12, 24, 36, 48, 60, 72, and 96 hpi. Eltanexor treatment show distinct effects on the transcript levels of three IE genes (UL36, UL122, and UL123). Eltanexor treatment increases the transcript levels of UL36 after 36 hpi, and appears not affect the transcript levels of UL123 throughout the first replication cycle (within 72 hpi), but reduce the transcript levels of UL122 after 36 hpi. Additionally, the transcript levels of E (UL44) and L (UL82, UL83, UL99, and UL75) genes are inhibited after 36 hpi (**Figure 3A**), compared with DMSO controls. To further determine whether Eltanexor treatment suppresses viral protein expression, Western-blotting assays were performed. The IE1 does not appear to be suppressed, but IE2 is significantly reduced by Eltanexor in comparison to DMSO controls. The protein

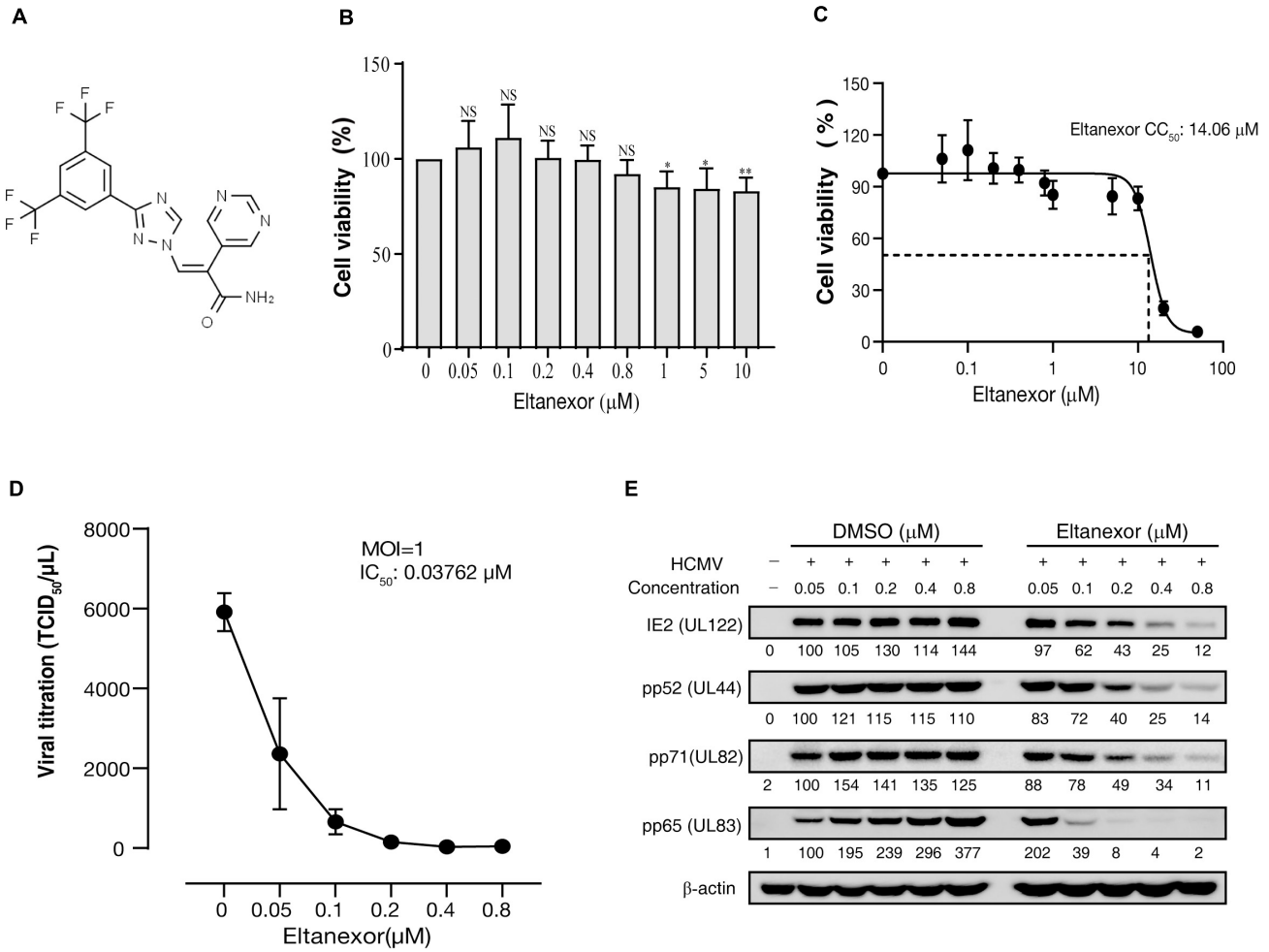


FIGURE 1 | Eltanexor inhibits HCMV lytic replication in HFFs in a dose-dependent manner. **(A)** Structure of Eltanexor (KPT-8602). **(B)** Effects of Eltanexor on cell viability were assayed at 72 h post treatment (hpt). HFFs were treated with Eltanexor at indicated concentrations or vehicle (DMSO, 0 μM). Cell viability was analyzed by a MTS-based colorimetric assays at 72 hpt. Data is presented as% of cell viability relative to vehicle control (indicated by 0 μM). Values represent mean ± SEM; $n = 5$. Statistical analyses were performed between indicated concentrations and 0 μM. $^{*}p < 0.05$, $^{**}p < 0.01$. **(C)** 50% cytotoxic concentration (CC50) is determined at 14.06 mM based on the results of cell viability assays according to non-linear trajectory analysis using GraphPad Prism. **(D)** Eltanexor inhibits the production of infectious virions in a dose-dependent manner. HFFs were infected with HCMV at a MOI = 1 and treated with the indicated concentrations of Eltanexor. Culture supernatants were collected at 96 hpi and further titrated by a $TCID_{50}$ method. Values represent mean ± SEM; $n = 3$. **(E)** Eltanexor inhibits viral protein synthesis in a dose-dependent manner. Cells were infected with HCMV at a MOI = 1 and then treated with DMSO or Eltanexor at the indicated concentrations for 72 h. Cell lysates were collected and viral protein expression was assayed by immunoblot for the following proteins: IE2, pp52, pp71, and pp65. Intensity values of viral protein bands were analyzed by Amersham Imager 600 analysis software. The experiments were repeated three times.

levels of E protein pp52 (encoded by UL44), pp65 (UL83), pp71 (UL82), and L protein pp28 (UL99) are suppressed after 36 hpi (**Figure 3B**). These results are consistent with the transcript levels of these genes observed in **Figure 3A**. In summary, our results indicate that Eltanexor treatment suppresses the protein levels of IE2 as well as other early and late viral proteins. The major IE protein IE2, encoded by UL122, is a crucial activator of viral late genes (Marchini et al., 2001). IE1, encoded by UL123, primarily functions as a coactivator with IE2 (Ahn and Hayward, 2000). Both IE1 and IE2 are derived from a same transcript but from two different spliced isoforms (Stinski and Petrik, 2008). the transcription of IE1 is less affected by Eltanexor treatment than IE2, likely because it is processed prior to IE2. Whether IE2 splice processing is inhibited by Eltanexor at the early stage of infection remains to be further determined.

Eltanexor Treatment Targets XPO1 for Proteasome-Mediated Degradation in HCMV Infected HFFs

To investigate the mechanism of how Eltanexor treatment suppresses HCMV replication, we examined whether Eltanexor directly targets XPO1. The protein levels of XPO1 appears not apparently upregulated in HCMV infection at indicated time points (**Figure 4A**), unlike in leukemia and other cancers (Vercruysse et al., 2017; Azizian and Li, 2020). During the course

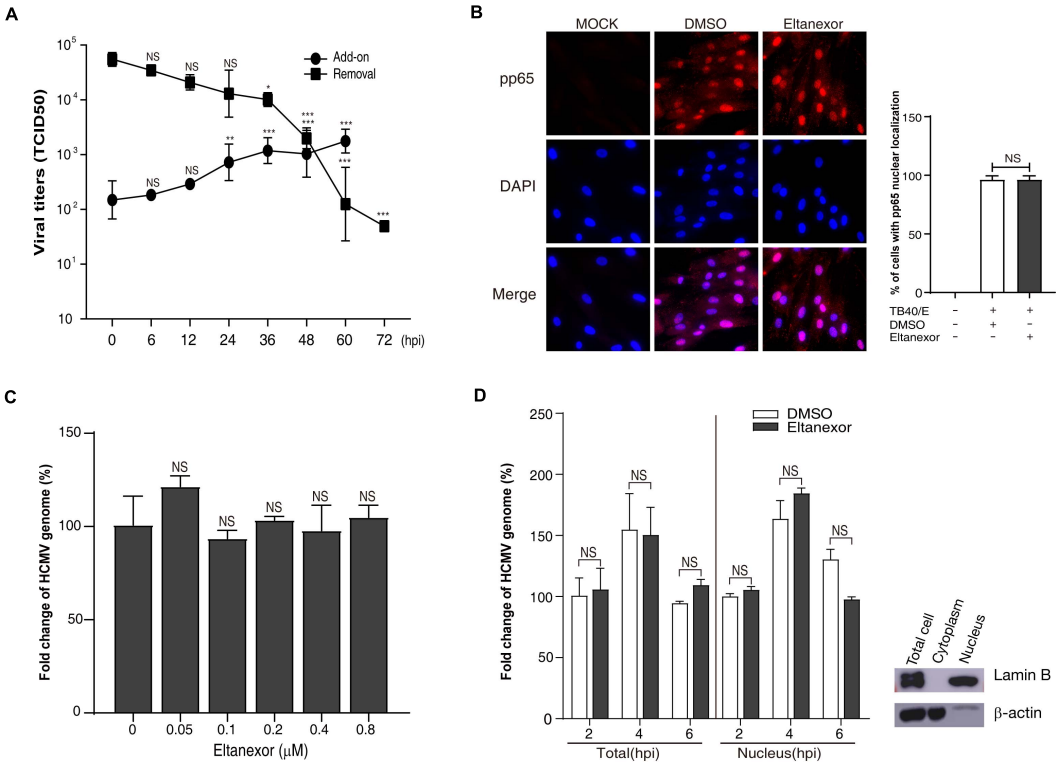


FIGURE 2 | Eltanexor treatment does not affect viral entry or nuclear trafficking. **(A)** Eltanexor add-on and removal assays showed viral replications was suppressed after 24 hpi. Culture supernatants collected from add-on or removal assays and HCMV titers were determined by a $TCID_{50}$ assay. Data represent mean ± SEM, $n = 3$. Statistical analyses were performed between indicated hpi and 0 hpi, $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$. **(B)** Nuclear localization of viral tegument protein pp65 is not affected by Eltanexor treatment. HFFs were infected with HCMV at a MOI = 1 and treated with Eltanexor (0.4 μM) or DMSO. At 6 hpi, cells were fixed and immunostained with mouse antibodies against pp65, followed by Cy3-conjugated goat anti-mouse antibody (red). Nuclei were stained with DAPI (blue). Percentages of cells with pp65 nuclear localization were calculated based on counts from 5 random fields with at least three hundred cells. The experiments were repeated 3 times. Data represent mean ± SEM, $n = 3$. NS stands for not significant. **(C)** Entry of the viral genome is not affected by Eltanexor treatment. HFFs were infected with HCMV at a MOI = 1 and treated with Eltanexor at different concentrations or DMSO for 2 h. Viral genomic DNA was isolated and quantified by qPCR with primers for UL122. Fold changes of viral genome in HCMV infected cells treated with 0 μM of Eltanexor were normalized to 100%. Statistical analyses were performed between different concentrations of Eltanexor and vehicle (0 mM). Data represent mean ± SEM, $n = 3$. **(D)** Nuclear trafficking of viral genomic DNA is not affected by Eltanexor treatment. HFFs were infected with HCMV at a MOI = 1 and then treated with Eltanexor (0.4 μM) or DMSO. Viral DNA isolated from total cells or nuclei at 2, 4, and 6 hpi were quantified by qPCR with primers for UL122. Fold changes of viral genome in HCMV infected cells treated with DMSO at 2 hpi were normalized to 100%. Data represent mean ± SEM, $n = 3$. Statistical analyses were performed between Eltanexor and DMSO treatments. Total cell lysates, cytoplasm and nuclear components were isolated with commercial cytoplasm-nuclear isolation kits and purity was confirmed by immunoblot against the nuclear marker Lamin B and the cytoplasmic marker β-actin. The experiments were repeated three times.

of HCMV infection (0∼96 hpi), the degradation of XPO1 was observed after 6 h post treatment with Eltanexor, but not in vehicle (DMSO)-treated control cells (**Figure 4A**). Eltanexor treatment decreases XPO1 levels, which is not dose-dependent (**Figure 4B**). When Eltanexor-treated cells is additionally treated with the proteasome inhibitor MG132 (5μM), the degradation of XPO1 is completely inhibited (**Figure 4C**). These results suggest that Eltanexor treatment inhibits HCMV replication by targeting XPO1 for degradation in a proteasome-dependent manner.

Eltanexor Treatment Increases the Nuclear Accumulation of Interferon Regulatory Factor 3 (IRF-3)

Upon HCMV infection, cytoplasmic IRF-3 is phosphorylated and dimerized, and then translocated from the cytoplasm into the nucleus (DeFilippis et al., 2006). The localization of IRF-3 is mediated by both a nuclear localization sequence (NLS) and a nuclear export sequence (NES) that are recognized by distinct shuttling receptors. The NLS and NES in IRF-3 are both constitutively active, but nuclear export is normally dominant. IRF-3 is exported from the nucleus to the cytoplasm via XPO1 (Kumar et al., 2000). Since Eltanexor treatment induces XPO1 degradation (**Figures 4A–C**), we speculate that it might increase nuclear localization of IRF-3 in HCMV infected cells. To investigate this hypothesis, we examined the localization of IRF-3 by immunofluorescence assays in HCMV infected HFFs treated with Eltanexor or DMSO (**Figure 5A**). When examined at 12, 24, 48, and 72 hpi, Eltanexor treatment increases the percentage of HCMV infected cells with nuclear localization of IRF-3 (**Figure 5B**). HCMV infection causes about 30% of cells to exhibit IRF3 nuclear localization, while Eltanexor treatment

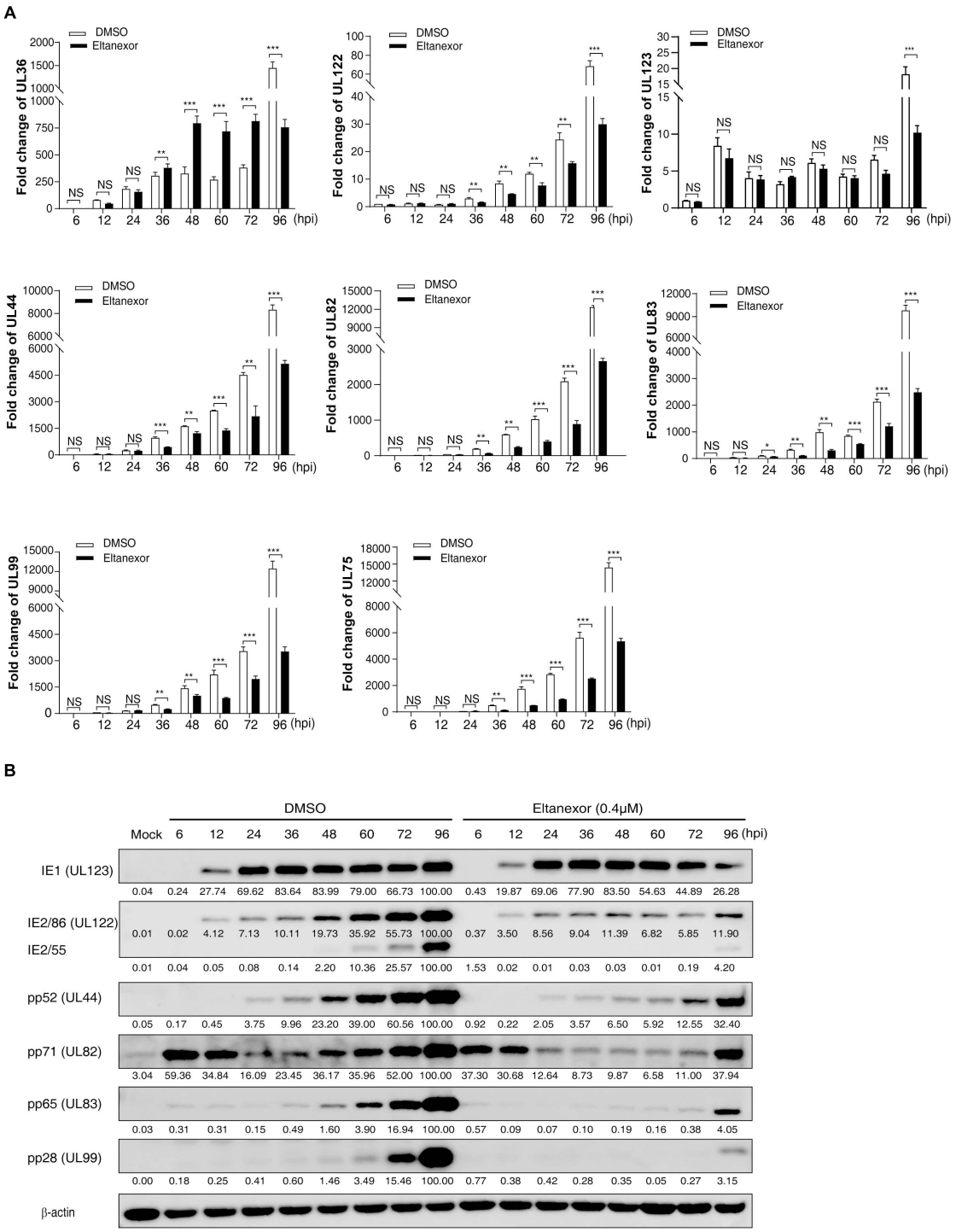


FIGURE 3 | Transcript and protein levels of HCMV are suppressed by Eltanexor treatment in HFFs. **(A)** Transcript levels of HCMV IE genes (UL36, UL122, and UL123), E gene (UL44), and L genes (UL82, UL83, UL99 and UL75) were quantified by RT-qPCR at the indicated time points. HFFs were infected with HCMV at a MOI = 1 and then treated with Eltanexor (0.4 μM) or DMSO. The levels of viral transcripts at the indicated time points were determined by RT-qPCR. Fold changes of viral transcripts were normalized to the internal control β-actin. Fold changes of viral transcripts in HCMV infected cells at 6 hpi were further normalized to 1 (arbitrary unit). Data represent mean ± SEM, $n = 3$, $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$. **(B)** protein levels of IE, E, and L genes were measured by Western-blotting at the indicated time points. Cell lysates were collected at the indicated time points as described in (A) and immunoblotted with antibodies against viral IE proteins (IE1 and IE2, encoded by UL123 and UL122), viral E protein (pp52, encoded by UL44), and viral L proteins (pp71, pp65, and pp28, encoded by UL82, UL83, and UL99). β-action served as the loading control. Viral protein band intensities were analyzed by Amersham Imager 600 analysis software. The experiments were repeated three times.

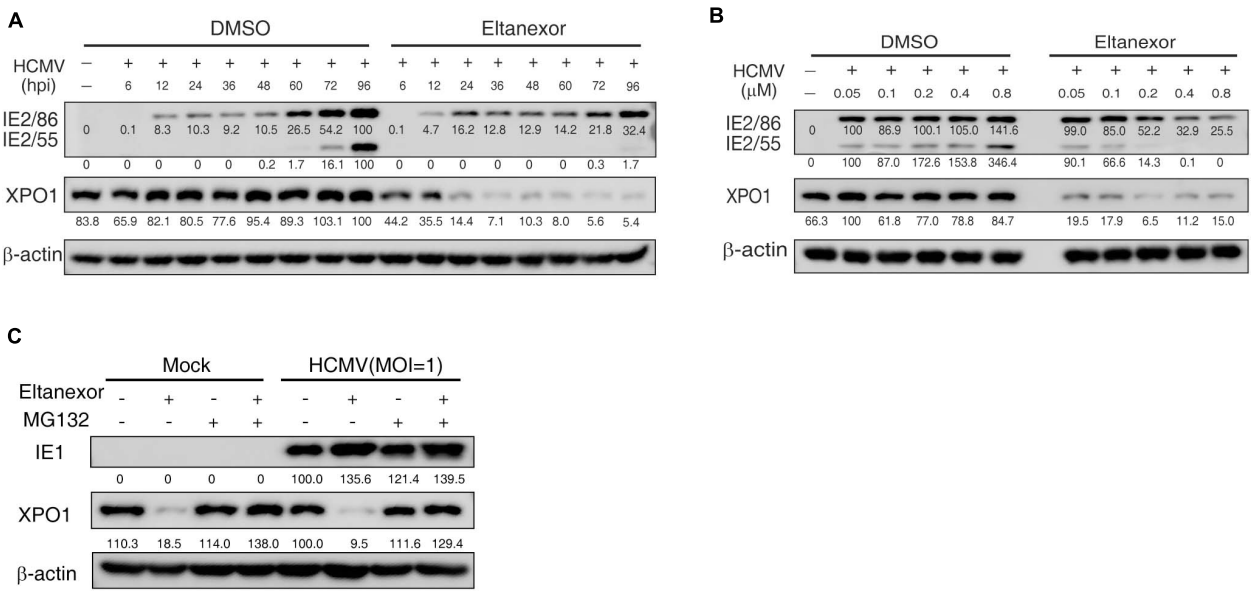


FIGURE 4 | Eltanexor treatment suppresses viral protein synthesis by targeting XPO1 for proteasome-mediated degradation. **(A)** Eltanexor treatment decreases XPO1 protein levels during the course of HCMV infection. HFFs were infected with HCMV at a MOI = 1 and treated with Eltanexor (0.4 μM) or DMSO. Cell lysates were collected at the indicated time points (hpi) and assayed by immunoblot for IE2, XPO1, and β-action. Protein band intensities were analyzed by Amersham Imager 600 analysis software. **(B)** Eltanexor treatment decreases the protein levels of XPO1 in a dose-independent manner. HFFs were infected with HCMV at MOI = 1 and treated with Eltanexor at the indicated concentrations or equal volumes of vehicle (DMSO). Cell lysate were collected at 72 hpi and immunoblotted and analyzed as in **(A)**. **(C)** Eltanexor induced XPO1 degradation is inhibited by the proteasome inhibitor MG-132. HFFs were infected with HCMV at a MOI = 1 on ice for 1.5 h, and then grown in complete medium with Eltanexor (0.4 μM) and/or the proteasome inhibitor MG132 (5μM). Cell lysates were collected and assayed by immunoblot for IE1, XPO1, and β-actin.

alone leads to nearly 100% cells with nuclear localization of IRF3. Viral dsRNA is a potent toll-like receptor 3 (TLR3) agonist that stimulates nuclear translocation of phosphorylated IRF3 and induces a type I interferon (IFN) response (Kumar et al., 2006; Murshid et al., 2016). Eltanexor treatment (0.4 μM) also increased the nuclear localization of IRF-3 in cells treated with a substitute dsRNA, polyinosine-polycytidylic acid [poly(I:C)] (**Figures 5C,D**), similar to that in HCMV-infected cells. In addition, Eltanexor treatment also inhibits the nuclear export of viral tegument protein pp65 which also contains a NES (**Figure 5E**), which was in line with a previously published study (Sanchez et al., 2007). Our results suggest that Eltanexor treatment increases nuclear retention of cellular IRF-3.

Eltanexor Treatment Increases Type I Interferon (IFN-I) Production in HCMV Infected HFFs

We next examined whether the nuclear accumulation of IRF3 induced by Eltanexor treatment could translate into increased production of type I IFN. Our results indicate that Eltanexor treatment (0.4 μM) increases the transcript levels of IFN-β in HCMV infected HFFs but not in uninfected cells at the indicated time points (**Figure 6A**). Similarly, Eltanexor treatment increases transcript levels of IFN-β in poly(I:C) treated cells but not in untreated cells (**Figure 6B**). However, Eltanexor alone in mock infected cells does not increase transcript levels of IFN-β, which is not consistent with the nuclear accumulation of IRF3 as observed in **Figure 5**. The protein levels of IFN-β in culture supernatants of HCMV infected or poly (I:C) treated cells determined by ELISA (**Figure 6C**) are in line with transcript levels of IFN-β (**Figures 6A,B**). Furthermore, transcript levels of interferon-stimulated genes ISG15 and ISG54 are also increased by Eltanexor treatment in HCMV infected cells at indicated time points (**Figure 6D**). Eltanexor alone could not stimulate IFN-β production, despite the nuclear accumulation of IRF3 as observed in **Figure 5**, which probably is due to lack of other transcriptional activators to form transcription initiation complexes, and thus is not sufficient to initiate the transcription of IFN-β. These results indicate Eltanexor increases the type I interferon response to inhibit HCMV replication.

DISCUSSION

The infection cycle of many viruses relies on XPO1 for transporting viral and cellular components from the nucleus to the cytoplasm (Mathew and Ghildyal, 2017). Interruption of nuclear export by selective inhibitors of nuclear export (SINEs) results in the reduction of viral protein expression and viral DNA replication, incomplete viral assembly, and proinflammatory responses (Elton et al., 2001; Pasdeloup et al., 2005; Cao and Liu, 2007; Sanchez et al., 2007; Ghildyal et al., 2009; Nakano and Watanabe, 2016). Eltanexor, as a second-generation inhibitor of XPO1, has been demonstrated to substantially reduce brain

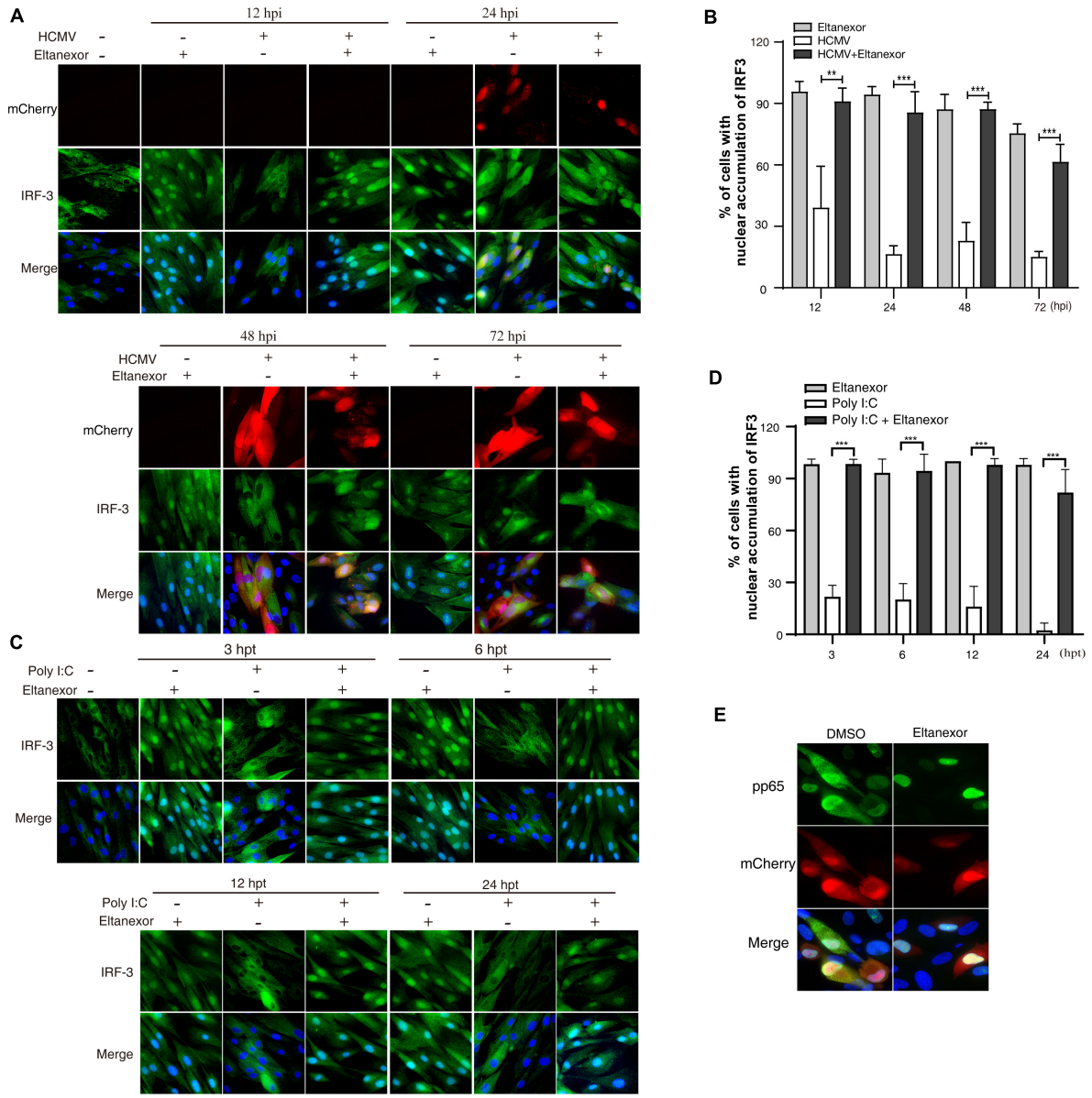


FIGURE 5 | Eltanexor treatment increases nuclear retention of IRF3 in HCMV infected or Poly (I:C) treated HFFs. **(A)** HFFs were infected with HCMV at a MOI = 1 and treated with Eltanexor (0.4 μM) or vehicle DMSO. At different time points, HFFs were fixed and immunostained with rabbit anti-IRF-3 antibody followed by FITC-conjugated goat anti-rabbit antibody (green). Nuclei were stained with DAPI (Blue). HCMV infected cells were indicated by mCherry (red). **(B)** Percentages of cells showing nuclear accumulation of IRF-3 in DMSO or Eltanexor treated HCMV-infected cells in **(A)** were calculated based on 5 random fields with more than 300 cells. The experiments were repeated 3 times. Data represent mean ± SEM, $n = 3$, $^{**}p < 0.01$, $^{***}p < 0.001$. **(C)** Eltanexor treatment increases the nuclear accumulation of IRF-3 in poly (I:C) treated HFFs. Poly (I:C) was added the culture medium at a concentration of 20μg/mL. HFFs were fixed and immunostained for IRF-3 as described in **(A)**. **(D)** Percentages of cells showing nuclear accumulation of IRF-3 in **(C)** were calculated based on 5 random fields. Data represent mean ± SEM; $n = 5$, $^{***}p < 0.001$. **(E)** Nuclear retention of viral tegument protein pp65 in Eltanexor treated HFFs at 72 hpi. HFFs were infected with HCMV at a MOI = 1, and treated with Eltanexor (0.4 μM) or vehicle DMSO. Cells were immunostained for pp65 (green). Nuclei were stained with DAPI (blue). HCMV infected cells were indicated by mCherry (red). The experiments were repeated 3 times.

penetration of hematopoietic malignancies with minimal toxicity to normal hematopoietic stem and progenitor cells compared to other SINEs (Das et al., 2015). Our study shows Eltanexor inhibits HCMV replication with better efficiency (EC_{50} = 0.03762 μM, CC_{50} = 14.06 μM, SI = 374) than Verdinexor (EC50 = 2.5 μM, CC_{50} = 73.3 μM, SI = 29) (Widman et al., 2018). In the preclinical study of Eltanexor (KPT8602) treating acute myeloid leukemia (Hing et al., 2016), the mean plasma concentrations reached 3 μM in primates orally dosed at 10 mg/kg, which is way higher than EC50 (0.03762 μM) for HCMV, suggesting Eltanexor might be suitable for anti-HCMV clinical trial in primates. The limitation of this study is that only single recombinant

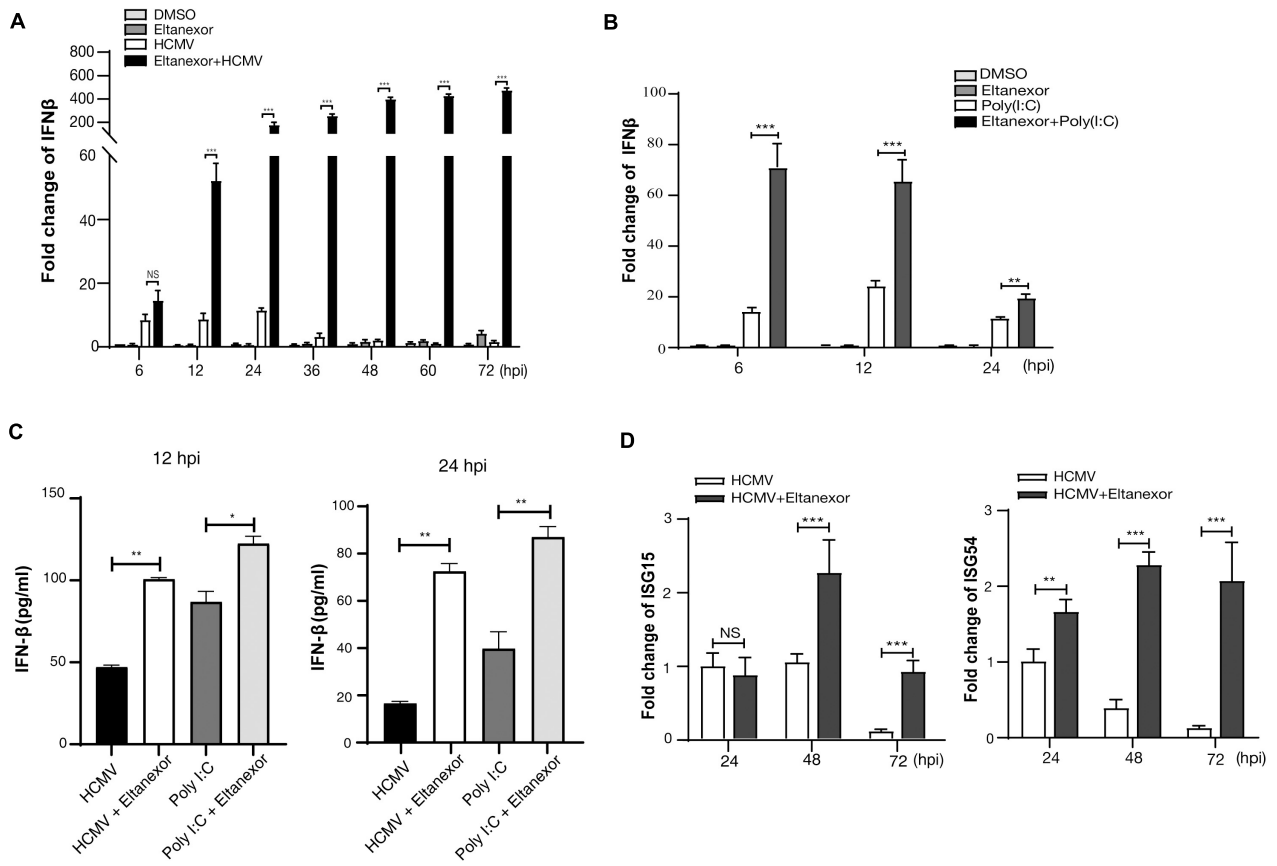


FIGURE 6 | Eltanexor treatment enhances a type I IFN response in HCMV-infected or poly (I:C) treated HFFs. **(A)** Transcriptional levels of IFN-β in HCMV infected cells treated with Eltanexor or DMSO were determined by RT-qPCR. HFFs were infected with HCMV at a MOI = 1, and treated with Eltanexor (0.4 μM) or DMSO. IFN-β mRNA levels were quantified by RT-qPCR at the indicated timepoints and normalized to cellular β-actin. Fold changes of IFN-β in DMSO-treated samples at 6 hpi were normalized to 1 (arbitrary unit). Data represent mean ± SEM, $n = 3$. **(B)** Eltanexor increase the transcriptional levels of IFN-β in poly(I:C) treated cells. HFFs were treated with DMSO, Eltanexor (0.4 μM), poly(I:C) (20μg/mL), or Eltanexor and poly(I:C). mRNA levels of IFN-β were quantified by RT-qPCR at the indicated timepoints and normalized to cellular β-actin. Data represent mean ± SEM, $n = 3$. **(C)** The production of IFN-β in the culture supernatants from HCMV infected or Poly (I:C) treated HFFs was determined by ELISA at the indicated timepoints. Data represent mean ± SEM, $n = 3$. **(D)** Expression of ISG15 and ISG54 in HCMV infected HFFs with or without Eltanexor treatment. Cells were infected with HCMV (MOI = 1) and treated with Eltanexor (0.4 μM) or vehicle DMSO. Transcript levels of ISG15 and ISG54 were determined by RT-qPCR at the indicated timepoints. Fold changes of transcripts in HCMV infected cells at 24 hpi were normalized to 1 (arbitrary unit). Data represent mean ± SEM, $n = 3$. $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.

HCMV strain was used for examining EC50, CC50, and SI. The index for resistant HCMV strains and other viruses can vary and further tests are warranted. Mechanistically, we have found that Eltanexor reduces the transcript and protein levels of IE2 (UL122) and other early and late genes post entry at the early stage of infection (within 24–36 hpi). In addition, Eltanexor treatment induces proteasome-mediated degradation of XPO-1 (**Figure 4**), increases the nuclear retention of IRF3 (**Figure 5**), and ultimately promotes production of IFN-β (**Figure 6**), which represents a novel antiviral mechanism of Eltanexor.

XPO1 overexpression is a common feature among many human cancer types, including lymphoma pancreatic, ovarian, glioma, lung, gastric, prostate, and colorectal cancers, and is associated with poor prognosis (Azizian and Li, 2020). Although suppression of XPO1-mediated nuclear export presents a unique therapeutic strategy, side effects of XPO1 inhibitors remain unclear. On the other hand, viral infection does not affect the expression levels of XPO1 as observed in this study (**Figures 4A,B**). The antiviral mechanism of SINEs is not fully understood, presumably by interrupting nuclear export of XPO1's cargo proteins at various stages of the viral lifecycle (Mathew and Ghildyal, 2017). Eltanexor directly inhibits XPO1-mediated nuclear export of viral proteins carrying a NES. HCMV replication is highly dependent on XPO1 for exporting both viral and cellular proteins of the nucleus (Sanchez et al., 2007; Frankenberg et al., 2012; Liu et al., 2012). Previous studies show the prototype XPO-1 inhibitor LMB impairs nuclear shuttling of viral proteins pp65 (Sanchez et al., 2007; Frankenberg et al., 2012). At the late phase of the lytic infection cycle, pp65 utilizes XPO1-mediated transport to shuttle from the nucleus to the cytoplasm, and disruption of XPO1-medited export of pp65 by LMB results in its nuclear retention, and dampens viral replication (Sanchez et al., 2007). Tegument protein UL94 is another late protein which relies on XPO1 mediated nuclear

export (Liu et al., 2012). Other tegument proteins such as pp71, UL35, UL47, and UL48, structural proteins pp150, pp28, and pIRS1/pTRS1, and non-structural protein UL97 are also translocated from the nucleus to the cytoplasm and play crucial roles in transcription of the viral genome, translocation of maturing nucleocapsids from the nucleus to the cytoplasm, modifying host cell response to infection, interfering with host cell cycle and optimizing the nuclear environment (Kalejta, 2008). XPO1 mediated nuclear export of viral proteins mentioned above normally occurs at the late stages of viral infection cycle (after 48 hpi). However, our results from Eltanexor add-on and removal assays show that Eltanextor treatment can significantly inhibit HCMV replication at the early stages of infection (within 24–36 hpi) (**Figure 2**). In particular, the expression of IE2 (UL122) is inhibited by Eltanexor treatment (**Figure 3**) at early times of infection, which can impede the progression of the rest of viral replication lifecycle. Thus, we speculate that Eltanexor exerts antiviral effects through targeting cellular or viral proteins that might impede the splicing or transcription of IE2 at the immediate-early and early stages of HCMV replication.

A recent deep proteomic study identified more than 1,000 proteins exported by XPO1 (Kirli et al., 2015), which means antiviral mechanisms of Eltanexor are complicated and might impact many signaling pathways. Eltanexor was reported to block cell proliferation and growth transformation by inducing p53-mediated cell cycle arrest in Kaposi's sarcoma-associated herpesvirus-transformed cells (Gruffaz et al., 2019). In addition, Yang et al. (2014) reported XPO1 inhibitors induced nuclear accumulation of TP53, leading to apoptosis of human melanoma cells. Hence, it is reasonable to postulate that Eltanexor treatment induce nuclear accumulation of p53 and enhance apoptosis in HCMV infected cells. Interestingly, upregulation of immediate early viral protein UL36 (**Figure 3A**) which encodes a viral inhibitor of caspase 8 induced apoptosis, might be a feedback response to counteract the pro-apoptotic effects of Eltanexor.

In this study, we found that Eltanexor treatment directly induced proteasome mediated degradation of XPO1 (**Figures 4A–C**) in HCMV infected cells. These results were consistent with XPO1 degradation in Eltanexor treated chronic lymphocytic leukemia cells (Hing et al., 2016). While many cellular proteins utilize XPO1, of particular interest to HCMV infection is IRF3, which relies solely on XPO1 for its nuclear export (Kumar et al., 2000). Type I IFNs are the first line of innate immune defense against HCMV infection. HCMV infection recruits a transcriptional complex containing interferon regulatory factor 3 (IRF-3) and acetyl transferases [CREB-binding protein (CBP) and p300] to the nucleus and thus stimulates type I IFN production (DeFilippis et al., 2006). Our results indicate Eltanexor treatment leads to increased nuclear accumulation of IRF3 in HCMV infected cells (**Figures 5A,B**), which contributes to increased expression of IFNβ as well as ISG15 and ISG54 (**Figure 6**). Therefore, we hypothesize that increased IFN-I production represents a novel antiviral mechanism of Eltanexor at the early stage of infection. A recent study showed XPO1 inhibition induces retention of autophagy adaptor protein p62 (SQSTM1) in the nucleus, which enhances activation of TBK1 and IRF3, resulting in increased expression of innate immune-related genes including IRF7, ISG15, IFIT1, IFIT2, and IFIT3, leading to a reduction of KSHV lytic replication (Meng and Gao, 2021). Thus, we speculate Eltanexor has potential for blocking a broad spectrum of viral pathogens through promoting type I interferon response.

XPO1 is considered a reasonable target for cancer therapy because overexpressed XPO-1 in cancers causes nuclear mislocalization of tumor suppressors and other regulatory proteins (Wang and Liu, 2019). The mechanisms of how Eltanexor targets cancer and viral pathogens might be quite different, and cautions should be taken for the side effects of Eltanexor in animal experiments and clinical trials of anti-cancer and antiviral therapies.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

AUTHOR CONTRIBUTIONS

QQ: conceptualization, writing—review and editing, supervision, project administration, and funding acquisition. YL: investigation, data curation, and writing—original draft preparation. YL, XK, TD, and QQ: formal analysis. YL, XK, and TD: validation. YL and QQ: methodology. YL and XK: software. YL and TD: visualization. All authors contributed to the article and approved the submitted version.

FUNDING

This study was funded by grants (2017KTSCX067 and 2020KZDZX1084) to QQ from the Department of Education of Guangdong province of China, a grant (2018A030307061) to QQ from the Natural Scientific Foundation of Guangdong province of China, a grant (200114115876753) to QQ from Science and Technology Bureau of Shantou City, and a grant (82072292) to QQ from the National Natural Science Foundation of China.

ACKNOWLEDGMENTS

We would like to thank Prof. Thomas Shenk at Princeton University for generously providing us with monoclonal antibodies against IE1/2 and Prof. Eain A. Murphy at SUNY Upstate Medical University for providing us with the HCMV BAC TB40/E*wt*-mCherry. We thank Prof. Shou-Jiang Gao at UPMC Hillman Cancer Center, University of Pittsburgh for instrumental scientific guidance. We also thank Prof. Cathy L. Miller at Iowa State University College of Veterinary Medicine, and Kate L. Carroll at the Department of Ophthalmology, University of Pittsburgh for proofreading our manuscript.

REFERENCES

Ahn, J. H., and Hayward, G. S. (2000). Disruption of PML-associated nuclear bodies by IE1 correlates with efficient early stages of viral gene expression and DNA replication in human cytomegalovirus infection. *Virology* 274, 39–55. doi: 10.1006/viro.2000.0448

Azizian, N. G., and Li, Y. (2020). XPO1-dependent nuclear export as a target for cancer therapy. *J. Hematol. Oncol* 13:61.

Azmi, A. S., Aboukameel, A., Bao, B., Sarkar, F. H., Philip, P. A., Kauffman, M., et al. (2013). Selective inhibitors of nuclear export block pancreatic cancer cell proliferation and reduce tumor growth in mice. *Gastroenterology* 144, 447–456. doi: 10.1053/j.gastro.2012.10.036

Boons, E., Vanstreels, E., Jacquemyn, M., Nogueira, T. C., Neggers, J. E., Vercruysse, T., et al. (2015). Human Exportin-1 is a target for combined Therapy of HIV and AIDS related lymphoma. *EBioMedicine* 2, 1102–1113. doi: 10.1016/j.ebiom.2015.07.041

Cannon, M. J. (2019). Congenital cytomegalovirus (CMV) epidemiology and awareness. *J. Clin. Virol.* 6(Suppl. 4), S6–S10.

Cao, Y., and Liu, X. Y. (2007). HIV-1 Rev and related inhibitors. *Yao Xue Xue Bao* 42, 347–351.

Cornell, R. F., Rossi, A. C., Baz, R., Hofmeister, C. C., and Kauffman, M. (2016). A Phase 1/2 study of the second generation selective inhibitor of nuclear export (SINE) compound, KPT-8602, in Patients with relapsed refractory multiple myeloma. *Blood* 128, 4509–4509. doi: 10.1182/blood.v128.22.4509.4509

Das, A., Wei, G., Parikh, K., and Liu, D. (2015). Selective inhibitors of nuclear export (SINE) in hematological malignancies. *Exp. Hematol. Oncol.* 4:7.

DeFilippis, V. R., Robinson, B., Keck, T. M., Hansen, S. G., Nelson, J. A., and Früh, K. J. (2006). Interferon regulatory factor 3 is necessary for induction of antiviral genes during human cytomegalovirus infection. *J. Virol.* 80, 1032–1037. doi: 10.1128/jvi.80.2.1032-1037.2006

Dong, X., Biswas, A., Suel, K. E., Jackson, L. K., Martinez, R., Gu, H., et al. (2009). Structural basis for leucine-rich nuclear export signal recognition by CRM1. *Nature* 458, 1136–1141. doi: 10.1038/nature07975

Douglas, C. M., Barnard, R., Holder, D., Leavitt, R., Levitan, D., Maguire, M., et al. (2020). Letermovir resistance analysis in a clinical trial of cytomegalovirus prophylaxis for hematopoietic stem cell transplant recipients. *J. Infect Dis.* 221, 1117–1126. doi: 10.1093/infdis/jiz577

Edward, S., Mocarski, J. T. S., Paul, D. G., and Pass, R. F. (2013). *Cytomegalovirus*, 6 Edn, Vol. 2. Philadelphia, PA: Wolters Kluwer Health/Lippincott Williams & Wilkins, c2013.

Elton, D., Simpson-Holley, M., Archer, K., Medcalf, L., Hallam, R., McCauley, J., et al. (2001). Interaction of the influenza virus nucleoprotein with the cellular CRM1-mediated nuclear export pathway. *J. Virol.* 75, 408–419. doi: 10.1128/jvi.75.1.408-419.2001

Fornerod, M., Ohno, M., Yoshida, M., and Mattaj, I. W. (1997). CRM1 is an export receptor for leucine-rich nuclear export signals. *Cell* 90, 1051–1060. doi: 10.1016/s0092-8674(00)80371-2

Frankenberg, N., Lischka, P., Pepperl-Klindworth, S., Stamminger, T., and Plachter, B. (2012). Nucleocytoplasmic shuttling and CRM1-dependent MHC class I peptide presentation of human cytomegalovirus pp65. *Med. Microbiol. Immunol.* 201, 567–579. doi: 10.1007/s00430-012-0269-7

Ghildyal, R., Ho, A., Dias, M., Soegiyono, L., Bardin, P. G., Tran, K. C., et al. (2009). The respiratory syncytial virus matrix protein possesses a Crm1-mediated nuclear export mechanism. *J. Virol.* 83, 5353–5362. doi: 10.1128/jvi.02374-08

Gruffaz, M., Yuan, H., Meng, W., Liu, H., Bae, S., Kim, J. S., et al. (2019). CRISPR-Cas9 screening of kaposi's sarcoma-associated herpesvirus-transformed cells identifies XPO1 as a vulnerable target of cancer cells. *mBio* 10:e00866-19.

Gutierrez, M., Brown, P. D. N., Gabrail, N., Baz, R., Reece, D. E., Savona, M., et al. (2013). Anti tumor activity of selinexor (KPT-330), a first-in-class oral selective inhibitor of nuclear export (SINE) XPO1/CRM1 antagonist in patients (pts) with relapsed / refractory multiple myeloma (MM) Or Waldenstrom's macroglobulinemia (WM). *Blood* 122, 1942–1942.

Hing, Z. A., Fung, H. Y., Ranganathan, P., Mitchell, S., El-Gamal, D., Woyach, J. A., et al. (2016). Next-generation XPO1 inhibitor shows improved efficacy and in vivo tolerability in hematological malignancies. *Leukemia* 30, 2364–2372. doi: 10.1038/leu.2016.136

Kalejta, R. F. (2008). Tegument proteins of human cytomegalovirus. *Microbiol. Mol. Biol. Rev.* 72, 249–265. doi: 10.1128/mmbr.00040-07

Kirli, K., Karaca, S., Dehne, H. J., Samwer, M., Pan, K. T., Lenz, C., et al. (2015). A deep proteomics perspective on CRM1-mediated nuclear export and nucleocytoplasmic partitioning. *Elife* 4:e11466.

Kosyna, F. K., and Depping, R. (2018). Controlling the gatekeeper: therapeutic targeting of nuclear transport. *Cells* 7:221. doi: 10.3390/cells7110221

Kudo, N., Matsumori, N., Taoka, H., Fujiwara, D., Schreiner, E. P., Wolff, B., et al. (1999). Leptomycin B inactivates CRM1/exportin 1 by covalent modification at a cysteine residue in the central conserved region. *Proc. Natl. Acad. Sci. U.S.A.* 96, 9112–9117. doi: 10.1073/pnas.96.16.9112

Kumar, A., Zhang, J., and Yu, F. S. (2006). Toll-like receptor 3 agonist poly(I:C)-induced antiviral response in human corneal epithelial cells. *Immunology* 117, 11–21. doi: 10.1111/j.1365-2567.2005.02258.x

Kumar, K. P., McBride, K. M., Weaver, B. K., Dingwall, C., and Reich, N. C. (2000). Regulated nuclear-cytoplasmic localization of interferon regulatory factor 3, a subunit of double-stranded RNA-activated factor 1. *Mol. Cell Biol.* 20, 4159–4168. doi: 10.1128/mcb.20.11.4159-4168.2000

Landolfo, S., Gariglio, M., Gribaudo, G., and Lembo, D. (2003). The human cytomegalovirus. *Pharmacol. Ther.* 98, 269–297.

Liu, Y., Zhang, Z., Zhao, X., Wei, H., Deng, J., Cui, Z., et al. (2012). Human cytomegalovirus UL94 is a nucleocytoplasmic shuttling protein containing two NLSs and one NES. *Virus Res.* 166, 31–42. doi: 10.1016/j.virusres.2012.02.023

London, C. A., Bernabe, L. F., Barnard, S., Kisseberth, W. C., Borgatti, A., Henson, M., et al. (2014). Preclinical evaluation of the novel, orally bioavailable selective inhibitor of nuclear export (SINE) KPT-335 in spontaneous canine cancer: results of a phase I study. *PLoS One* 9:e87585. doi: 10.1371/journal.pone.0087585

Marchini, A., Liu, H., and Zhu, H. (2001). Human cytomegalovirus with IE-2 (UL122) deleted fails to express early lytic genes. *J. Virol.* 75, 1870–1878. doi: 10.1128/jvi.75.4.1870-1878.2001

Mathew, C., and Ghildyal, R. (2017). CRM1 inhibitors for antiviral therapy. *Front. Microbiol.* 8:1171.

Matsuura, Y. (2016). Mechanistic insights from structural analyses of ran-GTPase-driven nuclear export of proteins and RNAs. *J. Mol. Biol.* 428, 2025–2039. doi: 10.1016/j.jmb.2015.09.025

Meloni, B. P., Milani, D., Edwards, A. B., Anderton, R. S., O'Hare Doig, R. L., Fitzgerald, M., et al. (2015). Neuroprotective peptides fused to arginine-rich cell penetrating peptides: neuroprotective mechanism likely mediated by peptide endocytic properties. *Pharmacol. Ther.* 153, 36–54. doi: 10.1016/j.pharmthera.2015.06.002

Meng, W., and Gao, S. J. (2021). Targeting XPO1 enhances innate immune response and inhibits KSHV lytic replication during primary infection by nuclear stabilization of the p62 autophagy adaptor protein. *Cell Death Dis.* 12:29.

Monecke, T., Güttler, T., Neumann, P., Dickmanns, A., Görlich, D., and Ficner, R. (2009). Crystal structure of the nuclear export receptor CRM1 in complex with Snurportin1 and RanGTP. *Science* 324, 1087–1091. doi: 10.1126/science.1173388

Murshid, A., Borges, T. J., Lang, B. J., and Calderwood, S. K. (2016). The Scavenger receptor SREC-I cooperates with toll-like receptors to trigger inflammatory innate immune responses. *Front. Immunol.* 7:226.

Nakano, K., and Watanabe, T. (2016). HTLV-1 rex tunes the cellular environment favorable for viral replication. *Viruses* 8:58. doi: 10.3390/v8030058

O'Connor, C. M., and Shenk, T. (2011). Human cytomegalovirus pUS27 G protein-coupled receptor homologue is required for efficient spread by the extracellular route but not for direct cell-to-cell spread. *J. Virol.* 85, 3700–3707. doi: 10.1128/jvi.02442-10

Pasdeloup, D., Poisson, N., Raux, H., Gaudin, Y., Ruigrok, R. W., and Blondel, D. (2005). Nucleocytoplasmic shuttling of the rabies virus P protein requires a nuclear localization signal and a CRM1-dependent nuclear export signal. *Virology* 334, 284–293. doi: 10.1016/j.virol.2005.02.005

Perwitasari, O., Johnson, S., Yan, X., Howerth, E., Shacham, S., Landesman, Y., et al. (2014). Verdinexor, a novel selective inhibitor of nuclear export, reduces influenza a virus replication in vitro and in vivo. *J. Virol.* 88, 10228–10243. doi: 10.1128/jvi.01774-14

Petosa, C., Schoehn, G., Askjaer, P., Bauer, U., Moulin, M., Steuerwald, U., et al. (2004). Architecture of CRM1/Exportin1 suggests how cooperativity is achieved during formation of a nuclear export complex. *Mol. Cell* 16, 761–775. doi: 10.1016/j.molcel.2004.11.018

Poole, C. L., and James, S. H. (2018). Antiviral therapies for herpesviruses: current agents and new directions. *Clin. Ther.* 40, 1282–1298. doi: 10.1016/j.clinthera.2018.07.006

Ranganathan, P., Yu, X., Na, C., Santhanam, R., Shacham, S., Kauffman, M., et al. (2012). Preclinical activity of a novel CRM1 inhibitor in acute myeloid leukemia. *Blood* 120, 1765–1773. doi: 10.1182/blood-2012-04-423160

Razonable, R. R. (2018). Drug-resistant cytomegalovirus: clinical implications of specific mutations. *Curr. Opin. Organ. Transplant.* 23, 388–394. doi: 10.1097/mot.0000000000000541

Sanchez, V., Mahr, J. A., Orazio, N. I., and Spector, D. H. (2007). Nuclear export of the human cytomegalovirus tegument protein pp65 requires cyclin-dependent kinase activity and the Crm1 exporter. *J. Virol.* 81, 11730–11736. doi: 10.1128/jvi.02760-06

Sinzger, C., Hahn, G., Digel, M., Katona, R., Sampaio, K. L., Messerle, M., et al. (2008). Cloning and sequencing of a highly productive, endotheliotropic virus strain derived from human cytomegalovirus TB40/E. *J. Gen. Virol.* 89, 359–368. doi: 10.1099/vir.0.83286-0

Stinski, M. F., and Petrik, D. T. (2008). Functional roles of the human cytomegalovirus essential IE86 protein. *Curr. Top. Microbiol. Immunol.* 325, 133–152. doi: 10.1007/978-3-540-77349-8_8

Sun, Q., Carrasco, Y. P., Hu, Y., Guo, X., Mirzaei, H., Macmillan, J., et al. (2013). Nuclear export inhibition through covalent conjugation and hydrolysis of Leptomycin B by CRM1. *Proc. Natl. Acad. Sci. U.S.A.* 110, 1303–1308. doi: 10.1073/pnas.1217203110

Turner, J. G., Dawson, J., Cubitt, C. L., Baz, R., and Sullivan, D. M. (2014). Inhibition of CRM1-dependent nuclear export sensitizes malignant cells to cytotoxic and targeted agents. *Semin. Cancer Biol.* 27, 62–73. doi: 10.1016/j.semcancer.2014.03.001

Vercruysse, T., De Bie, J., Neggers, J. E., Jacquemyn, M., Vanstreels, E., Schmid-Burgk, J. L., et al. (2017). The second-generation exportin-1 inhibitor KPT-8602 demonstrates potent activity against acute lymphoblastic leukemia. *Clin. Cancer Res.* 23, 2528–2541. doi: 10.1158/1078-0432.ccr-16-1580

Wang, A. Y., and Liu, H. (2019). The past, present, and future of CRM1/XPO1 inhibitors. *Stem. Cell Investig.* 6:6. doi: 10.21037/sci.2019.02.03

Widman, D. G., Gornisiewicz, S., Shacham, S., and Tamir, S. (2018). In vitro toxicity and efficacy of verdinexor, an exportin 1 inhibitor, on opportunistic viruses affecting immunocompromised individuals. *PLoS One* 13:e0200043. doi: 10.1371/journal.pone.0200043

Xu, D., Grishin, N. V., and Chook, Y. M. (2012). NESdb: a database of NES-containing CRM1 cargoes. *Mol. Biol. Cell* 23, 3673–3676. doi: 10.1091/mbc.e12-01-0045

Yang, J., Bill, M. A., Young, G. S., La Perle, K., Landesman, Y., Shacham, S., et al. (2014). Novel small molecule XPO1/CRM1 inhibitors induce nuclear accumulation of TP53, phosphorylated MAPK and apoptosis in human melanoma cells. *PLoS One* 9:e102983. doi: 10.1371/journal.pone.0102983

Zhang, K., Wang, M., Tamayo, A. T., Shacham, S., Kauffman, M., Lee, J., et al. (2013). Novel selective inhibitors of nuclear export CRM1 antagonists for therapy in mantle cell lymphoma. *Exp. Hematol.* 41, 67–78. doi: 10.1016/j.exphem.2012.09.002

Zheng, Y., Gery, S., Sun, H., Shacham, S., Kauffman, M., and Koeffler, H. P. (2014). KPT-330 inhibitor of XPO1-mediated nuclear export has anti-proliferative activity in hepatocellular carcinoma. *Cancer Chemother. Pharmacol.* 74, 487–495. doi: 10.1007/s00280-014-2495-8

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Microbiology

ORIGINAL RESEARCH
published: 01 June 2021
doi: 10.3389/fmicb.2021.662762

Diverse Trajectories Drive the Expression of a Giant Virus in the Oomycete Plant Pathogen *Phytophthora parasitica*

Sihem Hannat[1,2], Pierre Pontarotti[1,2,3], Philippe Colson[1,2,4], Marie-Line Kuhn[5], Eric Galiana[5], Bernard La Scola[1,2], Sarah Aherfi[1,2,4] and Franck Panabières[5]**

[1] Institut Hospitalo-Universitaire Méditerranée Infection, Marseille, France, [2] MEPHI, Institut de Recherche pour le Développement, Aix-Marseille Université, Marseille, France, [3] CNRS SNC5039, Marseille, France, [4] Assistance Publique - Hôpitaux de Marseille, Marseille, France, [5] INRAE, Université Côte d'Azur, CNRS, ISA, Sophia Antipolis, France

OPEN ACCESS

Edited by:
Masaharu Takemura, Tokyo University of Science, Japan

Reviewed by:
Eugene I. Savenkov, Swedish University of Agricultural Sciences, Sweden
David D. Dunigan, University of Nebraska-Lincoln, United States

***Correspondence:**
Sarah Aherfi
sarah.aherfi@ap-hm.fr
Franck Panabières
franck.panabieres@inrae.fr

Specialty section:
This article was submitted to Virology, a section of the journal Frontiers in Microbiology

Received: *01 February 2021*
Accepted: *07 May 2021*
Published: *01 June 2021*

Citation:
Hannat S, Pontarotti P, Colson P, Kuhn M-L, Galiana E, La Scola B, Aherfi S and Panabières F (2021) Diverse Trajectories Drive the Expression of a Giant Virus in the Oomycete Plant Pathogen Phytophthora parasitica. Front. Microbiol. 12:662762. doi: 10.3389/fmicb.2021.662762

Giant viruses of amoebas, recently classified in the class Megaviricetes, are a group of viruses that can infect major eukaryotic lineages. We previously identified a set of giant virus sequences in the genome of *Phytophthora parasitica*, an oomycete and a devastating major plant pathogen. How viral insertions shape the structure and evolution of the invaded genomes is unclear, but it is known that the unprecedented functional potential of giant viruses is the result of an intense genetic interplay with their hosts. We previously identified a set of giant virus sequences in the genome of *P. parasitica*, an oomycete and a devastating major plant pathogen. Here, we show that viral pieces are found in a 550-kb locus and are organized in three main clusters. Viral sequences, namely RNA polymerases I and II and a major capsid protein, were identified, along with orphan sequences, as a hallmark of giant viruses insertions. Mining of public databases and phylogenetic reconstructions suggest an ancient association of oomycetes and giant viruses of amoeba, including faustoviruses, African swine fever virus (ASFV) and pandoraviruses, and that a single viral insertion occurred early in the evolutionary history of oomycetes prior to the *Phytophthora–Pythium* radiation, estimated at ∼80 million years ago. Functional annotation reveals that the viral insertions are located in a gene sparse region of the *Phytophthora* genome, characterized by a plethora of transposable elements (TEs), effectors and other genes potentially involved in virulence. Transcription of viral genes was investigated through analysis of RNA-Seq data and qPCR experiments. We show that most viral genes are not expressed, and that a variety of mechanisms, including deletions, TEs insertions and RNA interference may contribute to transcriptional repression. However, a gene coding a truncated copy of RNA polymerase II along a set of neighboring sequences have been shown to be expressed in a wide range of physiological conditions, including responses to stress. These results, which describe for the first time the endogenization of a giant virus in an oomycete, contribute to challenge our view of *Phytophthora* evolution.

Keywords: gene transfer, giant viruses, *Phytophthora parasitica*, integration, endogenization, NCLDV, oomycetes

INTRODUCTION

Giant viruses of amoebas (GV), recently classified in the class Megaviricetes as part of the phylum Nucleocytoviricota, dramatically changed our view of the viral world following the description in 2003 of the first one, *Acanthamoeba polyphaga* mimivirus (La Scola et al., 2003). Since this initial discovery, dozens of GV have been isolated from the environment, animals, plants, and unicellular organisms (Delaroque and Boland, 2008; Derelle et al., 2008; Raoult and Boyer, 2010; Pagnier et al., 2013; Maumus et al., 2014; Sharma et al., 2014; Reteno et al., 2015; Leonard et al., 2018), or reconstructed from human and environmental metagenomes (Hingamp et al., 2013; Schulz et al., 2020). Accumulation of GV sequences has shown that these viruses display a mosaic genome architecture, encompassing a significant proportion of homologous sequences from viruses, eukaryotes, bacteria, and archaea, which were probably acquired by lateral gene transfers (Raoult et al., 2004; Filée et al., 2008). Taken together, these genes constitute a particularly original repertoire of hundreds of sequences, totally unprecedented in viruses (Colson et al., 2011a), which is still expanding (Moniruzzaman et al., 2020a; Rozenberg et al., 2020).

It is highly likely that invasion by GV deeply impacts the biology of their hosts. They may have deleterious effects upon integration within critical chromosomal regions. On the other hand, the extraordinary complexity of genes harbored by GV, such as transporters to take up nutrients, fermentation and photosynthesis genes, suggests that they may also complement some functions of their host (Vardi et al., 2012; Moniruzzaman et al., 2020a). Acquisition of novel functions provided by viral insertions may thus have dramatic consequences in the case of pathogenic microorganisms which have to cope with harsh and rapidly evolving environments. In this context, we identified a set of viral sequences possibly originating from a member of the Asfarviridae clade in the genome of the oomycete plant pathogen *Phytophthora parasitica* (Sharma et al., 2014).

Oomycetes constitute a deep lineage of lower eukaryotes that encompass some of the most notorious plant pathogens worldwide (Kamoun et al., 2015). Long time considered as fungi on the basis of structural features, shared ecological niches and common virulence strategies (Savory et al., 2015), oomycetes are grouped with brown algae diatoms and other unicellular organisms among Stramenopiles (Derelle et al., 2016). They include saprophytic organisms as well as highly aggressive animal and plant pathogens, such as the potato late blight agent or downy mildews. Most plant diseases are caused by members of the genus *Phytophthora,* which encompasses >180 formal species (Yang et al., 2017). The molecular bases of acquisition and evolution of virulence of these pathogens are a highly dynamic component that mainly relies on a sophisticated arsenal of effectors, most frequently extracellular proteins that are able to penetrate within host cells to defeat natural defense responses and manipulate plant functions to achieve successful infection (Franceschetti et al., 2017). Effectors may be secreted in the apoplast or in the cytoplasm, prior to reaching their site of action, which may be the cytoplasm, the plastid or even the nucleus.

In the present study, we aimed at gaining insights into the interactions between GV and the *Phytophthora* host genome by addressing the following questions: (i) Did the viral invasion occur one or several times during *Phytophthora*, and to a wider extent, oomycete evolution? (ii) Is it possible to trace the ancestral GV and approximately date the invasion event? (iii) What is the impact of GV insertions on the expression of the surrounding genomic environment?

MATERIALS AND METHODS

Sequence Manipulation and Phylogenetic Analyses

The *P. parasitica* strain INRA-310 infected region was previously described on a single contig under the accession KI669605.1 (Sharma et al., 2014). This contig, designed contig 2.45, was used for subsequent searches. Predicted genes were ascribed to *Phytophthora* or GV genomes by best hit Blastp analyses against NR at GenBank, excluding *Phytophthora* sequences from the dataset, and by Blastp and tBlastn searches against the homemade database containing the GV genomes and their corresponding predicted proteins. Viral candidate sequences were retrieved in the oomycete genomes publicly available after tBlastn searches in the whole-genome contig subset of GenBank using default parameters. Conserved domains and putative functions of predicted proteins were searched using Blastp analyses on various databases, including PFAM (Mistry et al., 2020), Interpro (Blum et al., 2020), SMART (Letunic et al., 2020), and Superfamily (Pandurangan et al., 2019). Transposable elements (TEs) were searched by BlastN searches against the RepBase database using the Censor tool (Bao et al., 2015). A tRNA search was conducted using both tRNAscan-SE (Lowe and Chan, 2016) and Aragorn (Laslett and Canback, 2004). Sequences were aligned by using MUSCLE software (Edgar, 2004). Date of viral integration was estimated using the TimeTree tool (Kumar et al., 2017), based on different molecular clock methods (Matari and Blair, 2014). Phylogenetic trees were built by using the maximum likelihood method, which was carried out with the FastTree program (Price et al., 2010), with 1000 bootstrap replications, using JTT template substitution (standard default).

Phytophthora Strains, Growth Media

Phytophthora parasitica strain INRA-310 was maintained in the INRAE ISA collection and grown on V8 medium at 24°C. *P. parasitica* mycelial cultures were submitted to diverse abiotic stresses as follows: initially, axenic cultures were grown at 37°C instead of 24°C to mimic heat shock. Alternatively, culture medium was supplemented with NaCl at a high concentration (0.6M), mimicking saline conditions encountered in drought. Finally, mycelial cultures were submitted to sub-lethal doses of $CuSO_4$ (at a final 0.3M concentration). Initially employed as the famous Bordeaux mixture, this compound is commonly used in conventional agriculture to control oomycete diseases, at the increasing expense of environmental safety (Erwin and Ribeiro, 1996). Stress treatments were conducted for 4 h at 24°C. The mycelium was then harvested and thoroughly rinsed before

nucleic acid extraction. Following these various treatments, expression of viral genes was assessed by quantitative RT-PCR.

Expression Analyses

Expression of viral genes was assessed according two strategies. First, we estimated the total amount of reads matching each ORF of the contig in Blastn searches against RNA-Seq data available at GenBank (Accessions SRX1124837–SRX1124840, SRX1124842–SRX1124845, SRX1124847–SRX1124868, SRX27 27839–SRX2727852, SRX4902085–SRX4902107). In parallel, genes of interest were analyzed by qRT-PCR. For the RT-PCR, genomic DNA was extracted from 10 day-old cultures conducted on V8 medium as previously described (Panabières and Le Berre, 1999). RNA was extracted by using the RNeasy Mini Kit (QIAGEN, Germany). After the addition of RNAse, the DNA was digested by TURBO DNase (Invitrogen Thermo Fisher Scientific, Lithuania) at 37°C, three times in 30 min. Real-time RT-PCR was performed using two-step, one-step experiments.

For the two steps RT-PCR, cDNA was synthesized from 1 μg of RNA, using the SuperScript^{TM} VILO^{TM} cDNA Synthesis Kit Invitrogen (Thermo Fisher Scientific) reaction mixture following the manufacturer's instructions. Gene expression was assessed by qPCR with gene-specific primers (Eurogentec, United States; **Supplementary Table 1**) and the fluorescent dye SYBR-Green (Invitrogen, Vilnius, Lithuania). The PCR amplification program consisted of: 95°C for 5 min following 45 cycles of 95°C for 10s, 59°C for 20s, and 72°C for 30s. Gene expression was considered efficient if there was an amplification curve with a Ct value $\leq$35 in each triplicate experiment with a melting temperature identical to those obtained on the DNA extracts (**Table 3**).

The one-step RT PCR reaction mixture consisted of RNA, using the SuperScript kit^{TM} III RT/Platinum^{TM} according to the manufacturer's instructions. All assays were performed in triplicate and included negative controls (DNA/RNA-free PCR mix). A gene of *P. parasitica* strain INRA 310 encoding for the 40S ribosomal protein S3A (WS21, Genbank accession number: XM_008905737.1), known to be transcribed in all tested conditions, was used as an internal control (Yan and Liou, 2006).

Sanger Sequencing

The PCR amplicon generated from purified cDNA of *P. parasitica* was sequenced on the Applied Biosystems 3130xl Genetic Analyzer (Thermo Fisher Scientific, France) using the BigDye Terminator DNA Sequencing Kit (Perkin-Elmer) according to the manufacturer's instructions. The sequence was assembled using ChromasPro 2.0.0 software and compared to the NCBI database by Blastn analysis.

RESULTS

Integration of GV Sequences in Oomycete Genomes

In the ∼550-kb contig from *P. parasitica* INRA-310 (contig 2.45, GenBank accession KI669605) 17 predicted genes matched with GV. When excluding the hits from *Phytophthora,* a total of 12 had GV hits as the best match (Sharma et al., 2014). This finding predated the achievement of several genome sequencing projects and the availability of an increasing number of complete genomes of oomycetes, including tens of *Phytophthora* genomes. Similarly, an invaluable number of studies on GV was conducted that totally modified our view of the viral world, expanding the phylogenetic tree of GV to thousands of members (Schulz et al., 2020). Therefore, we first performed a new gene calling step, prior to conducting sequential best Blast hit searches against public databases, excluding *P. parasitica*, then *Phytophthora*, and finally oomycetes from the datasets.

Gene calling revealed 7 additional ORFs of 180–1,005 nucleotides (60–335 codons) that were added to the ORFs already predicted in the contig, so that a total of 126 putative proteins were used as queries in the best Blast hit search, among which 15 appeared to better match GV sequences (**Supplementary Table 2**). These sequences mostly corresponded to partial protein sequences when compared to the best viral homologs, and notably encoded RNA polymerases I and II, the major capsid protein, and a DEAD helicase (**Supplementary Table 2**). We noted that the RNA polymerase II was represented by several ORFs, indicating a likely duplication or several transfer events (**Supplementary Figure 1**). The other viral sequences did not display any domain enabling functional annotation. No tRNA or ribosomal RNA genes were found across this contig.

Some viral candidate sequences were found to best match with oomycete sequences from *Phytophthora cactorum* or even *Pythium oligandrum*, whose genomes were released in public databases after the publication of the initial work (Armitage et al., 2018; Yang et al., 2018; Faure et al., 2020). This suggests a possible viral invasion of oomycete genomes having occurred before the *Phytophthora-Pythium* speciation. Alternatively, this might indicate that invasion events by GV occurred several times during oomycete evolution. Sequential best Blast hit searches did not reveal evidence of viral signatures that would be shared by all oomycetes. The viral candidates were therefore searched in oomycete genomes through tBlastn searches against the contig sequences hosted at the NCBI GenBank whole genome shotgun database. In line with independent reports (Yutin et al., 2014; Gallot-Lavallée and Blanc, 2017), we retrieved homologs of some of the viral candidate genes in several, but not all *P. parasitica* genomes publicly available and deposited under the denomination *P. parasitica* or the synonym *Phytophthora nicotianae* (Panabières et al., 2016). We also retrieved homologs in various *Phytophthora* species, as well as in at least three species formerly belonging to the *Pythium* genus, and recently ascribed to *Globisporangium,* among the order Pythiales (Uzuhashi et al., 2010), (**Table 1**). Therefore, the presence of GV sequences within the *P. parasitica* PPINRA-310 genome was not the outcome of a possible contamination, but rather reflected one or several viral insertion events affecting ancestral oomycetes.

The successive best Blast hit searches also revealed genes of ambiguous origin for which the 10 first hits corresponded to various unrelated organisms. Among them, we identified genes encoding a putative histone H3 (PPTG_14870), a DNA

TABLE 1 | Distribution of the viral candidates from PPINRA-310 among oomycetes. Indicated are the species name and contig accession.

ORF number	Description	*Phytophthora*	Accession	*Pythium*	Accession
PPTG_14861	RNA Polymerase 2	*P. parasitica*		*Globisporangium irregulare*	NCVO01008180.1
		P. quercina	JACBOW010000075.1	*Globisporangium ultimum*	ADOS01001492.1
		P. castanetorum	JACBOV010000007.1	*Pythium oligandrum*	LSAJ01000108.1
		P. ramorum	RHLB01000784.1	*Globisporangium iwayamae*	AKYA02010070.1
		P. boehmariae	JAAVTJ010000040.1		
		P. constricta	JAAVTI010000058.1		
PPTG_14866	Major Capsid Protein	*P. parasitica*		*Globisporangium irregulare*	NCVO01008180.1
		P. cactorum	*NBIJ01015056.1*	*Globisporangium ultimum*	ADOS01001492.1
		P. cryptogea	*AUWJ02017648.1*		
		P. constricta	*JAAVTI010000058.1*		
		P. quercina	JACBOW010000075.1		
		P. x alni	*AUPN01065155.1*		
		P. cambivora	*AUVH01118500.1*		
PPTG_23628	T5orf172 domain protein	*P. parasitica*			
		P. fragariae	*QXGF01000047.1*		
PPTG_14881	RNA Polymerase 1	*P. parasitica*		*Globisporangium irregulare*	AKXZ02009300.1
		P. cactorum	*NBIJ01015056.1*		
		P. syringae	*JAAKBD010000413.1*		
		P. vignae	*JABJXB010000139.1*		
		P. taxon totara	*LGSO01000030.1*		
		P. quercina	JACBOW010000075.1		
		P. x alni	*AUPN01065155.1*		
		P. cambivora	*AUVH01118500.1*		
PPTG_14885	RNA Polymerase 1	*P. parasitica*		*Globisporangium irregulare*	NCVO01006854.1
		P. rubi	*QXFV01000589.1*		
		P. cactorum	*NBIJ01015056.1*		
		P. x alni	*AUPN01065155.1*		
		P. cambivora	*AUVH01118500.1*		
		P. taxon totara	*LGSO01000030.1*		
PPTG_14890	DEAD-like helicase	*P. parasitica*			
		P. pinifolia	*AWVW02037847.1*		
		P. cactorum	*NBIJ01015056.1*		
		P. castanetorum	JACBOV010000007.1		
		P. ohioensis	*JACBOX010000020.1*		
		P. quercina	JACBOW010000075.1		
		P. pinifolia	*AWVW02037847.1*		
		P. cambivora	*AUVH01118500.1*		
PPTG_14893	Hypothetical protein	*P. parasitica*		*Pythium periplocum*	MRVE01000072.1
		P. cactorum	*NBIJ01015056.1*	*Globisporangium ultimum*	AKYB02034954.1
				Pythium oligandrum	NAJK01000032.1
gene 787	Hypothetical protein	*P. parasitica*			
PPTG_14900	Hypothetical protein	*P. parasitica*			
gene 788	Hypothetical protein	*P. parasitica*			
gene 789	Hypothetical protein	*P. parasitica*			
PPTG_14924	Hypothetical protein	*P. parasitica*			
PPTG_14926	RNA Polymerase 2	*P. parasitica*		*Globisporangium irregulare*	NCVO01008180.1
				Pythium oligandrum	SPLM01000008.1
PPTG_14927	RNA Polymerase 2	*P. parasitica*		*Globisporangium irregulare*	NCVO01008180.1
				Globisporangium ultimum	ADOS01001492.1
		P. ramorum	RHLB01000784.1	*Pythium oligandrum*	LSAJ01000108.1
gene 404	Hypothetical protein	*P. parasitica*			
		P. ohioensis	JACBOX010000017.1		

primase (PPTG_23627) and two partial ORFs (PPTG_14947 and PPTG_14951) corresponding to a ribonucleotide reductase. Such genes had already been predicted in several GV genomes (Derelle et al., 2008; Colson and Raoult, 2010; Yoshikawa et al., 2019). In addition, the histone gene was found to be restricted to *P. parasitica* and strikingly derived from canonical histones from oomycetes. Further phylogenetic analyses are necessary to precisely conclude whether this sequence is of viral origin or if it was acquired through lateral gene transfer involving other organisms.

Invasion by an *Asfarviridae* Member Likely Occurred Once During Oomycete Evolution

Sequences of unambiguous viral origin were not located randomly across the contig 2.45, but rather accumulated at 3 discrete regions that spanned the whole contig length (**Figure 1A**). We thus observed that the two blocks encoding RNA polymerase II (PPTG_14861 and PPTG_14926/PPTG_14927p) were separated by >335 kb. Despite their overall conservation (78% identity, 85% positive over a 147-amino acids region, **Supplementary Figure 1**), these two genes might have originated from distinct viral donors and therefore from successive invasion events. In addition, looking at the best Blast hits indicated that the well-defined viral sequences found in *P. parasitica* (RNA Pol, Capsid, helicase), located in the blocks I and II, had better affinities to genes from asfarviruses, including Pacmanvirus, African swine fever virus (ASFV), and faustoviruses, while hypothetical proteins located at the block III better matched various viruses, with markedly less confident e-values. We then conducted a series of maximum likelihood phylogenetic analyses with a focus on RNA polymerases and the major capsid protein (MCP) to know whether the *P. parasitica* genome was invaded once by a single virus (or members of a single clade) or in successive events. Phylogenetic reconstructions showed that the viral sequences found in the genomes from oomycetes constitute a robust clade closely related to *Asfarviridae* (**Figures 2**, **3**). Phylogenetic reconstruction based on the RNA polymerase II and the capsid protein suggested that the viral invasion occurred prior to the speciation between *Phytophthora* and *Pythium* (**Figures 3** and **Supplementary Figure 2**). We included a sequence from RNA polymerase of viral origin found in the genome of *Hyphochytrium catenoides*, a basal organism belonging to Stramenopiles (Leonard et al., 2018). Several genes likely acquired from an *Asfarviridae* donor were identified within the genome of *H. catenoides*, among which were RNA polymerases and MCP (Gallot-Lavallée and Blanc, 2017; Leonard et al., 2018). The tree topologies indicate that *Hyphochytrium* more likely acquired viral genes from an asfarvirus (**Figures 2**, **3** and **Supplementary Figure 2**). The exact nature of the GV donor that invaded the *Phytophthora*/*Pythium* lineages is uncertain, but it clearly belonged to the group that includes asfarviruses and faustoviruses, which are relatives (Reteno et al., 2015). We could roughly estimate the date of integration of this donor into the genome of the *Phytophthora*/*Pythium* ancestor to 80 million years ago.

Viral Sequences Are Imbedded Among Transposable Elements Within a Gene Sparse Region of the *P. parasitica* Genome

We intended to assign a function to genes present on the contig 2.45 in addition to likely viral sequences. Predicted proteins were generally of relatively small length (mean = 219, median = 166 amino acids). Identification of functional domains was achieved in only a few cases, so that the majority of ORFs encoded hypothetical proteins. Annotation revealed sequences relevant to protein-protein interactions, such as ankyrin repeat proteins, IPT/TIG domain proteins and chaperones, or components of the mRNA turnover and silencing. We also identified four members of the CRN (Crinkler and Necrosis) effector superfamily (Franceschetti et al., 2017), as well as several remnants of TEs (**Supplementary Table 2**). This observation prompted us to screen the whole contig for the presence of TEs and repetitive sequences. This task was difficult, because the *P. parasitica* strain INRA-310 genome was assembled from Illumina-derived data. Repetitive sequences including TEs are known to frequently escape identification in genomes. Indeed, their repetitive nature and their size, longer than the average read length generated by the Illumina technique, prevent their identification and lead to collapsing them in most genome assemblies (Panabières et al., 2020). Consequently, a substantial proportion of the contig length was made of N-stretches of unassembled regions, introducing numerous gaps in the assembly. Nonetheless, a search against RepBase revealed ∼40 TE-derived sequences that were scattered all along the contig 2.45, as already observed with viral sequences (**Figure 1B**). Further analysis revealed that retrotransposons were represented by 14 sequences, overwhelmingly corresponding to Gypsy-like elements, while class II transposons were mainly represented by Polintons, PiggyBac, MuDR, and Helitrons (**Table 2**). The paucity of protein-coding sequences associated to a wide abundance of TEs, which is itself highly underestimated because of the number of N-stretches in the contig, indicated that the viral insertion occurs in a gene sparse region.

The Viral Locus Retained a Discrete Transcriptional Activity

To evaluate the potential impact of the viral sequences on *P. parasitica* biology, we intended to assess the expression of viral candidates in a variety of physiological situations. To this end, we first collected RNA-Seq data generated on *P. parasitica* that are publicly available. We mined libraries prepared from various pre-infection stages of *P. parasitica*, including libraries enriched in small RNAs. Overall, the contig was transcriptionally silent, although ∼20 ORFs organized into 10 short loci were expressed to various extent, as estimated by the total number of reads obtained in Blastn searches (**Supplementary Table 3**). Genes displaying some expression levels encoded CRNs, as well as proteases and hypothetical proteins, and a predicted tyrosine recombinase (PPTG_14945), possibly reflecting the activity of a TE. Sequences located in the block I were not expressed,

FIGURE 1 | location of the viral candidate ORFs and identified TEs on the 550-kb contig. For clarity, *Phytophthora* predicted genes have not been integrated in the figure. **(A)** Viral candidates. Only ORFs with a known putative function are numbered. **(B)** Location of TE-derived domains identified using the Censor tool at Repbase. DNA transposons are represented by green blocks and retrotransposons are represented by red rectangles.

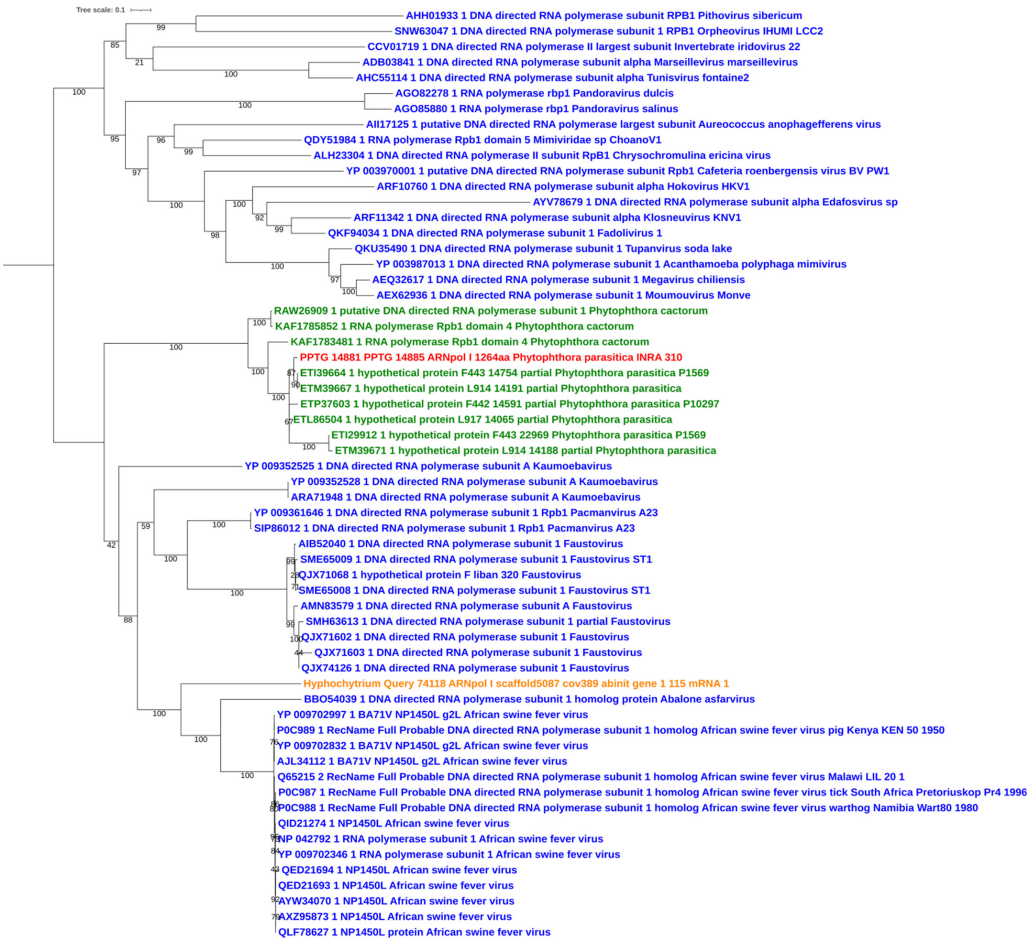


FIGURE 2 | Maximum likelihood tree of RNA polymerase I amino acid sequences. The alignment was built using MUSCLE, and the iTol visualization was used with the rooted midpoint option. Sequences from NCLDVs are indicated in blue and sequences from oomycetes are indicated in green. The viral sequence from *P. parasitica* strain INRA-310 is indicted in red, and the viral sequence identified in the genome of *Hyphochytrium catenoides* (see text) is shown in orange. Bootstrap values are given below nodes in percent.

while some ORFs from the blocks II and III were expressed to various extent. A refined analysis of RNA-Seq data from ORFs of the block I revealed that the RNA pol II-encoding gene PPTG_14861, as well as sequences corresponding to RNA pol I, were slightly expressed as small RNAs, which constitute potent non-coding RNA regulators (**Supplementary Table 3**). On the other hand, sequences from the block II encoding RNA pol II (represented by the partial sequences PPTG_14926 and PPTG_14927) were expressed in all situations observed. They were embedded among a set of genes that are also expressed in an apparent constitutive manner. Interestingly, this gene was located inside a set of genes that were also expressed. So, RNA

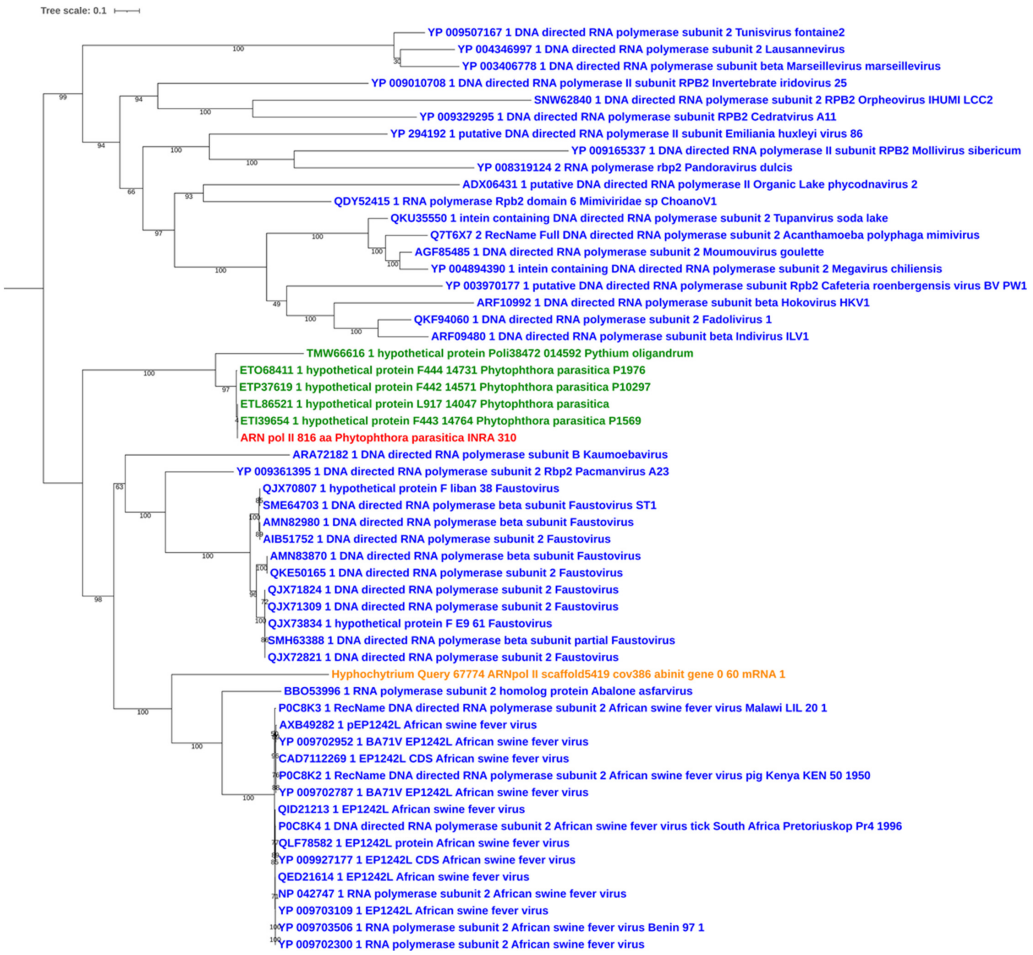


FIGURE 3 | Maximum likelihood tree of RNA polymerase II protein. The alignment was built using MUSCLE, and the iTol visualization was used with the rooted midpoint option. Sequences from NCLDVs are indicated in blue and sequences from oomycetes are indicated in green. The viral sequence from *P. parasitica* strain INRA-310 is indicted in red, and the viral sequence identified in the genome of *Hyphochytrium* catenoides (see text) is shown in orange. Bootstrap values are given below nodes in percent.

pol II appeared to be the sole gene of viral origin being expressed under the conditions used for the generation of RNA-Seq data. We then investigated the potential expression of viral genes under various physiological conditions never explored to date. To this aim, qRT-PCR experiments by both SYBR Green and hydrolysis probe real-time methodologies were used.

Only six ORFs, namely PTG_14900, PPTG_14924, PPTG_14926, PPTG_14927, PPTG_14866, and PPTG_14884, appeared to be expressed using SYBR Green methodology in the control condition. Moreover, we assessed gene transcription under stress conditions consisting in a heat shock, NaCl and copper sulfate. As these stress experiments were known to mimic environmental conditions and to modify the level of expression of several *Phytophthora* genes, we aimed to assess these conditions may also affect the expression of the genes that have homologs with GV. Each of the six genes appeared to be expressed in at least one of the three stress conditions (**Table 3**). Notably, a positive result was found for PPTG_14927 in the 3 stress conditions, including heat shock, which usually leads to gene repression. Given the relatively low specificity of the SYBR green methodology, and intending to confirm these first results, we tentatively amplified these ORFs using specific hydrolysis probes. Interestingly, two of these six ORFs, PPTG_14926 and PPTG_14927, provided a significant expression signal in each situation, and displayed strongly reproducible results (**Table 4**). Further Sanger sequencing of the PCR products confirmed that the sequences amplified derived from the PPTG_14926 and PPTG_14927 targets. These experiments thus partially confirmed exploitation of RNA-Seq data, as no evidence of transcription was found for most genes, with the noticeable exception of PPTG_14926 and PPTG_14927, corresponding to a region of RNA polymerase II.

DISCUSSION

The present study constitutes an additional step for the analysis conducted a few years ago that revealed the presence of GV

TABLE 2 | Distribution of Transposable elements across the contig 2.45.

Gene Id	Length (bp)	start	end
PiggyBac	642	18,925	19,567
PiggyBac	3,689	29,772	33,461
Copia	2,152	37,086	39,238
Polinton	494	68,331	68,825
Gypsy	1,509	79,718	81,227
Gypsy	2,082	86,958	89,040
Gypsy	13,700	113,775	127,475
MuDR	2,367	135,354	137,721
Polinton	164	179,461	179,625
Copia	6,247	179,840	186,087
Polinton	521	220,225	220,746
PiggyBac	6,299	223,052	229,351
Polinton	1,491	230,030	231,521
Polinton	1,720	240,788	242,508
Gypsy	3,298	242,878	246,176
Polinton	2,022	248,763	250,785
Polinton	6,016	303,155	309,171
PiggyBac	829	311,637	312,466
Gypsy	8,860	313,262	322,122
ISL2EU	908	335,963	336,871
Polinton	686	340,477	341,163
Gypsy	3,820	357,908	361,728
Gypsy	2,145	373,639	375,784
Gypsy	770	378,581	379,351
Polinton	268	382,048	382,316
Polinton	580	411,058	411,638
Helitron	248	415,117	415,365
Gypsy	506	419,285	419,791
MuDR	546	421,183	421,729
Gypsy	2,096	437,585	439,681
Mariner	483	448,659	449,142
Polinton	231	453,381	453,612
Gypsy	2,634	504,928	507,562
Gypsy	4,753	514,061	518,814
Polinton	397	518,992	519,389
Helitron	2,160	524,094	526,254
MuDR	556	528,211	528,767
Polinton	4,503	533,253	537,756
L1	2,985	541,466	544,451

TABLE 3 | Results of SYBR Green RT-PCR performed on the cDNA *P. parasitica* INRA-310 prepared from mycelial cultures grown under different abiotic stress conditions. (P: Positive; N: Negative; Non-stressful condition: Control, stressful conditions with: 37°C, NaCl, CuSo4).

ORFs	Control	37°C	NaCl	$CuSO_4$
PPTG_14900	P	N	N	P
PPTG_14924	P	N	P	P
PPTG_14926	P	N	N	P
PPTG_14884	P	N	P	N
PPTG_14866	P	N	N	P
PPTG_14927	P	N	P	P
Control (WS21)	P	P	P	P

sequences in the genome of the oomycete plant pathogen *P. parasitica*. Beyond this initial observation, we show here that viral insertion is a relatively ancient event, as it involved several *Phytophthora* species as well as members of *Globisporangium*, a new genus erected among the former *Pythium* genus (Uzuhashi et al., 2010). We first searched viral candidates among various *P. parasitica* genomes and confirmed their presence throughout the species, although to various extents. Such discontinuous distribution of viral sequences among *P. parasitica* isolates may have several explanations. First, viral insertions reside in a particularly dynamic region of the genome harboring TEs. Consequently, the majority of *P. parasitica* genomes that were sequenced according to the Illumina technique are poorly resolved at this locus, as illustrated by the substantial proportion of N-stretches found in the INRA-PP310 sequences or the small length of the contigs harboring viral sequences in the other genomes. Supporting this explanation is the finding that complete *P. parasitica* genomes sequenced by long read approaches, especially those of tobacco pathogens designated under the synonym *P. nicotianae* (Liu et al., 2016; Panabières et al., 2016), possess the viral insertions. Therefore, the lack of several ORFs, including the expressed region encoding RNA polymerase II, may result from differences in the sequencing methodology, and may also correspond to chromosomal deletions among strains. *P. parasitica* is considered to have a particularly broad host range, being able to infect more than 250 plant families (Panabières et al., 2016). And yet, frequent cases of host specificity are observed, and strains collected on tobacco have long been considered genetically distinct (Lacourt et al., 1994; Biasi et al., 2016). Specific searches relying on PCR-based identification conducted on a large range of *P. parasitica* strains, instead of Blast-based analyses conducted on a limited number of genomes available from *P. parasitica* strains, would be necessary prior to inferring relationships between the presence of viral insertions and the origin of the strains, with eventual consequences from these insertions on the virulence and host specificity of their recipient strains.

The discontinuous distribution of viral sequences that was observed among *Phytophthora* spp. is disconnected from the existing phylogeny of the genus *Phytophthora*, although *P. parasitica* shares taxonomical affinities among the phylogenetic clade 1 with *P. cactorum* (Yang et al., 2017) that appear to possess more conserved viral insertions than the other genomes analyzed. Finally, no trace of viral sequences was found outside *Phytophthora,* and in few *Pythium* members, when analyzing oomycetes. This observation is of interest, as the genus *Pythium* was for a long time considered a sister to the *Phytophthora* clade within a Pythiaceous lineage (Ascunce et al., 2017) before its recent redefinition into five genera (Uzuhashi et al., 2010). The conservation of viral sequences observed during the Blast searches therefore supports the hypothesis of a common ancestor being invaded by a single viral lineage shortly before the *Phytophthora-Pythium* radiation. In this evolutionary framework, the absence of any viral sequence in the known genomes of the *Hyaloperonospora* genus is worth noting. Indeed, this genus, represented by the obligate parasite *H. arabidopsidis*,

TABLE 4 | Results of Platinum™ RT-PCR performed on the RNA of *P. parasitica* INRA-310 prepared from mycelial cultures grown under different abiotic stress conditions. (N: Negative; Ct: Cycle threshold).

ORFs	Control			37°C			NaCl			$CuSO_4$		
	Ct-1	Ct-2	Ct-3	Ct-1	Ct-2	Ct-3	Ct-1	Ct-2	Ct-3	Ct-1	Ct-2	Ct-3
PPTG_14900	N	N	N	N	N	N	N	N	N	N	N	N
PPTG_14924	N	N	N	N	N	N	N	N	N	N	N	N
PPTG_14926	28.0	28.0	28.1	29.2	29.4	28.7	29.7	29.9	30.0	29.9	29.9	29.8
PPTG_14884	N	N	N	N	N	N	N	N	N	N	N	N
PPTG_14866	N	N	N	N	N	N	N	N	N	N	N	N
PPTG_14927	29.6	29.5	29.8	26.4	27	26	27.2	27.3	26.7	26.2	25.6	25.2
Control (WS21)	24.8	25.1	27.3	21.8	20.7	21.4	23.8	25.6	23.2	24.1	24.0	29.2

appears closer to *Phytophthora* than *Pythium* in the majority of phylogenomic analyses (Matari and Blair, 2014; Ascunce et al., 2017; Mccarthy and Fitzpatrick, 2017). Similarly, there was no evidence of viral sequences in the complete genomes of downy mildews that have been clustered with *Phytophthora* spp. in these phylogenomic reconstructions. One possibility is that the lifestyle of these organisms, characterized by obligate parasitism, created a particular ecological niche which protected them from further viral invasions.

Phylogenetic analyses indicated that oomycete genomes were infected by several members of the family Asfarviridae, among which are Pacmanvirus, ASFV, and faustoviruses. This contrasts with findings by Moniruzzaman et al. (2020b), who identified viral sequences most similar to mimiviruses or phycodnaviruses in 12 algal genomes. The relatively modest genome sequence conservation among the asfarvirus relatives found here clearly shows that asfarvirus relatives, isolated and sequenced during the last decades (Reteno et al., 2015; Andreani et al., 2017), largely diverged from their common ancestor, or that new members of this clade have still to be uncovered. Asfarvirus giant relatives were mainly characterized following co-cultivation with amoebas or after isolation on natural hosts (Boyer et al., 2009; Pagnier et al., 2013). Recent surveys of metagenomic data have dramatically expanded the diversity of GV in terms of amounts of clades, but have not expanded the intra-clade diversity of Asfarviridae (Schulz et al., 2018, 2020; Bäckström et al., 2019). Therefore, we propose that the viral insertions observed in *P. parasitica*, and to a lesser extent in other oomycetes, reflect the sequences of an Asfarviridae ancestor rather than a member(s) of this clade that remains to be discovered. Oomycetes belong to Stramenopiles, which includes several organisms that have also integrated pieces of GV (Delaroque and Boland, 2008; Gallot-Lavallée and Blanc, 2017; Leonard et al., 2018). Recent phylogenetic studies showed that these various organisms retained sequences from a single different viral host. Notably, the deep branching *Hyphochytrium* and oomycetes display viral sequences of asfarviruses, whereas other stramenopiles were rather infected by a phaeovirus that infects marine algae (Gallot-Lavallée and Blanc, 2017). We note that infected oomycetes (members of *Phytophthora* and *Pythium, sensu lato*) and *Hyphochytrium* members are terrestrial, flagellated organisms, whereas the other analyzed stramenopiles are all marine microorganisms. This suggests that GV display some host specificity that is driven by ecological constraints. The six viral genes identified during this study have been frequently found in other cases of genome invasions and are shared by numerous GVs. They include the conserved genes encoding RNA polymerases I and II, the major capsid protein, and a helicase. This is puzzling and questions the significance and role of these genes. They were not eliminated, which may suggest that they are useful. Other sequences of likely viral origin without any evidence of a given function, the abundance of which is a hallmark of these GV genomes, have been characterized (Boyer et al., 2010; Schulz et al., 2017; Yoshikawa et al., 2019).

Furthermore, several genes harbored by the contig are of ambiguous origin and may be of viral origin as well. Among them, we found a histone, a DNA primase, a ribonucleotide reductase, and a plethora of hypothetical proteins which are restricted to *P. parasitica*. Whether they were acquired during viral infection is unknown, although their lateral acquisition is likely. On the other hand, the contig hosted few protein-coding sequences of oomycete origin, among them effectors, and is populated by a substantial proportion of TEs. This characteristic is found in oomycete and fungal genomes and is known as "two-speed genome" architecture, with regions enriched in TEs and pathogenicity-related genes (effectors) alternating with other gene-rich regions (Dong et al., 2015; Seidl and Thomma, 2017). Three sequences of probable oomycete origin were present in several copies across the contig, which are ankyrin repeat proteins, IPT/TIG domain proteins and CRN effectors. Ankyrins mediate protein-protein interactions, and thus participate in multiple cellular processes (Li et al., 2006). They also contribute to animal immunity and plant defense against pathogens (Ji et al., 2016; Ngaki et al., 2016). Conversely, ankyrins also act as effectors in the obligate bacterium *Anaplasma phagocytophilum* to manipulate host chromatin and gene expression (Rennoll-Bankert et al., 2015). Therefore, ankyrins might be a component of the *Phytophthora* virulence arsenal. Ankyrin repeat-containing proteins are also particularly prevalent in giant viruses (Iyer et al., 2006; Colson and Raoult, 2010; Mirzakhanyan and Gershon, 2020). One possibility is that the ankyrin-encoding ORFs uncovered on the contig 2.45 were acquired during viral invasion. Whether hybrid ankyrin complexes of both viral and cellular origins are established upon viral infection is unknown. Furthermore, the apparent lack of expression of the ankyrins present in the contig under the experimental conditions explored makes the function of these proteins obscure. Like

ankyrins, IPT/TIG (immunoglobulin-like, plexins, transcription factors/transcription factor immunoglobulin) domains have been suggested to be involved in protein-protein interactions and DNA binding (Lara-Ramírez et al., 2017). Finally, CRN effectors constitute a large superfamily of proteins that have been identified across several kingdoms, but have been mainly studied in pathogenic oomycetes (Amaro et al., 2017). They are translocated into the host cells and contribute to virulence, and are thought to target the host nucleus and develop DNA-damaging activities (Stam et al., 2013; Camborde et al., 2019). Ankyrins, IPT/TIG proteins and CRNs may collectively fulfill related functions through targeting nuclear partners.

The contig analyzed in the present work also contained numerous repetitive sequences, among which are several known classes of TEs. This observation has important consequences. We may suppose that overall, TEs have been mobilized to interrupt the viral genes and protect the *Phytophthora* genome from the potential deleterious effects of viral invasion. Yet, the two main classes present in the contig, PiggyBac and Polintons, are frequently associated with viruses. PiggyBac was initially isolated from a baculovirus infecting the lepidopteran *Trichoplusia* (Gilbert et al., 2014), before further identification in a wide range of organisms, including oomycetes (Haas et al., 2009; Gaulin et al., 2018). This DNA transposon is also co-opted by herpesviruses, resulting in a new transposon entity that is able to infect fish genomes (Inoue et al., 2017). On the other hand, Polintons (aka Maveriks (Pritham et al., 2007) share several structural characteristics and an evolutionary ancestry with virophages, which are small (15–25 kb) dsDNA viruses that infect giant virus replication factories and may limit the replication of their viral host (La Scola et al., 2008; Mougari et al., 2019). Several analyses suggest that virophages are the progenitors of Polintons (Fischer and Suttle, 2011; Krupovic et al., 2014; Campbell et al., 2017). Polintons would then constitute a case of endogenization of viral elements. Therefore, the proportion of genes of viral origin in the contig 2.45 may be strikingly higher than initially proposed.

Using two different approaches, we show that at least one gene of viral origin is expressed in *P. parasitica* under a wide range of physiological conditions. Among the three stress conditions, only heat shock led to a decrease in genes. Hence, a copy of the RNA polymerase II is expressed at significant levels, as observed in qPCR experiments after analysis of RNA-Seq data. GV sequences have been previously identified in plant genomes, but they are transcriptionally silent (Maumus et al., 2014). However, traces of expression of viral sequences have been found in the transcriptome of the brown algae *Ectocarpus siliculosus* (Delaroque and Boland, 2008), another member of the Stramenopile lineage, like oomycetes. A close examination of expression data and of the architecture of the invaded contig reveals striking features. Hence, several distinct events may lead to inhibition of viral gene expression. First, interruption of the reading frames. This is clearly observed in the case of RNA Pol II (PPTG_14861), which only represents a truncated region of the viral homolog. We may suppose that the ancient, initial insertion of the viral gene was followed by successive mutations and rearrangements, likely driven by neighboring TEs. This agrees with the hypothesis of a massive mobilization of TEs that represents a large part of the contig, and which might have invaded it to inactivate the viral genes. Second, we observed that several viral genes are only transcribed in the form of small RNAs, although to a very moderate extent. It suggests that epigenetic mechanisms regulate their expression. There is compelling evidence that epigenetics, especially small RNA-based silencing, plays a major role in the biology of *Phytophthora*. Recent studies have shown that this mechanism targets protein coding genes like effectors and also TEs and other repetitive sequences like satellite DNA families (Vetukuri et al., 2012; Åsman et al., 2016; Panabières et al., 2020). Finally, viral genes may be silenced through fusion with TEs. Hence the gene encoding major capsid protein (PPTG_14866) is fused to a PiggyBac-derived sequence at its 3'end that likely contributes to its inactivation. In this context of global inactivation of viral insertions, the detection of the second RNA polymerase II gene using two non-redundant regions (PPTG_14926 and PPTG_14927) is enigmatic. Sequence alignment with the best viral hit revealed that the *P. parasitica* gene may correspond to a truncated copy that would be logically silent. In addition, we did not identify upstream sequences that would act as a potential promoter, but the 5' moiety of the gene was flanked by a sequence corresponding to Harbinger, a DDE DNA transposon (Wicker et al., 2007). And it is located in a region containing 9 expressed ORFs, while the rest of the contig is globally transcriptionally silent. We may thus suppose a beneficial effect of this local environment, and that its transcription is driven by surrounding sequences, among which the TE is a good, but not the only, candidate. Although we showed that these two ORFs derived from the RNA polymerase II were transcribed, it is not certain that they were translated into functional proteins. However, their RNA might be used as a defense mechanism, as previously described for the moss *Physcomitrella patens*. Indeed, a recent analysis of the *P. patens* genome suggests that giant viruses embedded in the genome were transcribed during gametogenesis and were used as siRNA-mediated silencing to protect the gametes from viral infection (Lang et al., 2018). Whatever the transcriptional status of the RNA polymerase II in *P. parasitica*, it would indicate that viral insertion was followed by an endogenization process. Whether the transcription of this truncated RNA polymerase gene provides an advantage to *Phytophthora* is unclear, but undoubted, as it is the sole gene of viral origin to have escaped the inactivation process.

CONCLUSION

This study constitutes the first description of a set of GV genes in the genome of an oomycete. The origins of these sequences largely remain obscure, but we here provide evidence that invasion of this class of major plant pathogens by a member of the family Asfarviridae was an ancient event. Our results raise intriguing questions about the relationship between TEs and viruses and possible shared steps in their evolutionary histories. These questions also pave the way for future studies aiming at a better knowledge of the nature and extent of the sequence flows that shaped *Phytophthora* genomes, with the evidence of

lateral transfers. Also, traces of expression of at least two genes of viral origin indicates a likely GV endogenization event, and the general question of the importance of endogenous viral sequences in the evolution of eukaryotic genomes constitutes an active field of research (Geering et al., 2014; Frank and Feschotte, 2017; Moniruzzaman et al., 2020b). Our data adds to previous knowledge indicating that the presence of genes encoding RNA polymerase subunits and/or major capsid protein from GV in eukaryotic organisms appears to be a general phenomenon. Indeed, this was observed for various eukaryotes that are demonstrated or possible hosts of GV (Colson et al., 2011b; Maumus et al., 2014; Moniruzzaman et al., 2020b). A widespread endogenization of GV sequences was recently reported in various green algae, which consisted of large endogenous viral elements, and more fragmented footprints of past GV integration were also observed (Moniruzzaman et al., 2020b). These giant endogenous viral elements exhibited gene duplications and losses and were associated with introns invasions and transposons. The role and significance of viral sequences in the evolution of eukaryotic genomes continues to expand our knowledge of viral invasion and of the mosaicism of eukaryotic genomes. This is particularly important in the case of pathogenic microorganisms, as in the case of oomycetes, which are prone to permanent adaptation to their environment. An important contribution to this research would be the development of proteomics approaches to validate hypotheses arising from gene expression studies.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: NCBI GenBank, accession no: NW_008634126.1, SRX1124837–SRX1124840, SRX1124842–SRX1124845, SRX1124847–SRX1124868, SRX27, 27839–SRX2727852, and SRX4902085–SRX4902107.

AUTHOR CONTRIBUTIONS

SA, BL, PP, and FP conceived the study. SH, M-LK, and EG conducted *Phytophthora* cultures and performed nucleic acids purification. FP and EG conducted manual annotation, data mining, and RNA-Seq analyses. SH and PP performed phylogenetic reconstructions. SH conducted qPCR analyses. SH, SA, and FP wrote the manuscript. BL and PC supervised the study. BL, PC, and EG revised the manuscript. All authors reviewed and approved the final manuscript.

FUNDING

This work was partially funded by INRAE and by the French Government (National Research Agency, ANR) through the "Investments for the Future" LabEx SIGNALIFE: program (reference #ANR-11-LABX-0028-01).

ACKNOWLEDGMENTS

We sincerely thank Djamal Brahim Belhaouari and Louis Tsakou-Ngouafo for their technical help.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2021.662762/full#supplementary-material

Supplementary Figure 1 | Alignment of RNA Polymerase 2 encoding ORFs.

Supplementary Figure 2 | Maximum likelihood tree of MCP. The figure legend is the same as in **Figure 2**.

Supplementary Table 1 | primers used in this study.

Supplementary Table 2 | Characteristics of predicted ORFs of the 550-kb contig from *P. parasitica* INRA-310 containing the 'viral locus'. Sequences of likely viral origin are indicated in bold characters.

Supplementary Table 3 | Relative expression of each ORF of the contig 2.45 as the total number of reads matching each sequence. *WS21* was used as internal constitutive control.

REFERENCES

Amaro, T. M. M. M., Thilliez, G. J. A., Motion, G. B., and Huitema, E. (2017). A perspective on CRN proteins in the genomics age: evolution, classification, delivery and function revisited. *Front. Plant Sci.* 8:99. doi: 10.3389/fpls.2017.00099

Andreani, J., Khalil, J. Y. B., Sevvana, M., Benamar, S., Di Pinto, F., Bitam, I., et al. (2017). Pacmanvirus, a new giant icosahedral virus at the crossroads between Asfarviridae and Faustoviruses. *J. Virol.* 91, e212–e217. doi: 10.1128/jvi.00212-17

Armitage, A. D., Lysøe, E., Nellist, C. F., Lewis, L. A., Cano, L. M., Harrison, R. J., et al. (2018). Bioinformatic characterisation of the effector repertoire of the strawberry pathogen *Phytophthora cactorum*. *PLoS One* 13:e0202305. doi: 10.1101/321141

Ascunce, M. S., Huguet-Tapia, J. C., Ortiz-Urquiza, A., Keyhani, N. O., Braun, E. L., and Goss, E. M. (2017). Phylogenomic analysis supports multiple instances of polyphyly in the oomycete peronosporalean lineage. *Mol. Phylogenet. Evol.* 114, 199–211. doi: 10.1016/j.ympev.2017.06.013

Åsman, A. K. M., Fogelqvist, J., Vetukuri, R. R., and Dixelius, C. (2016). *Phytophthora infestans* Argonaute 1 binds microRNA and small RNAs from effector genes and transposable elements. *New Phytol.* 211, 993–1007. doi: 10.1111/nph.13946

Bäckström, D., Yutin, N., Jørgensen, S. L., Dharamshi, J., Homa, F., Zaremba-Niedwiedzka, K., et al. (2019). Virus genomes from deep sea sediments expand the ocean megavirome and support independent origins of viral gigantism. *MBio* 10, 1–23. doi: 10.1128/mBio.02497-18

Bao, W., Kojima, K. K., and Kohany, O. (2015). Repbase update, a database of repetitive elements in eukaryotic genomes. *Mob. DNA* 6:11.

Biasi, A., Martin, F. N., Cacciola, S. O., Magnano Di San Lio, G. M., Grünwald, N. J., and Schena, L. (2016). Genetic analysis of *Phytophthora nicotianae* populations from different hosts using microsatellite markers. *Phytopathology* 106, 1006–1014. doi: 10.1094/PHYTO-11-15-0299-R

Blum, M., Chang, H.-Y., Chuguransky, S., Grego, T., Kandasaamy, S., Mitchell, A., et al. (2020). The InterPro protein families and domains database: 20 years on. *Nucleic Acids Res.* 49, 344–354. doi: 10.1093/nar/gkaa977

Boyer, M., Gimenez, G., Suzan-Monti, M., and Raoult, D. (2010). Classification and determination of possible origins of ORFans through analysis of nucleocytoplasmic large DNA viruses. *Intervirology* 53, 310–320. doi: 10.1159/000312916

Boyer, M., Yutin, N., Pagnier, I., Barrassi, L., Fournous, G., Espinosa, L., et al. (2009). Giant *Marseillevirus* highlights the role of amoebae as a melting pot in emergence of chimeric microorganisms. *Proc. Natl. Acad. Sci. U. S. A.* 106, 21848–21853. doi: 10.1073/pnas.0911354106

Camborde, L., Raynaud, C., Dumas, B., and Gaulin, E. (2019). DNA-Damaging effectors: new players in the effector arena. *Trends Plant Sci.* 24, 1094–1101. doi: 10.1016/j.tplants.2019.09.012

Campbell, S., Aswad, A., and Katzourakis, A. (2017). Disentangling the origins of virophages and polintons. *Curr. Opin. Virol.* 25, 59–65. doi: 10.1016/j.coviro.2017.07.011

Colson, P., Gimenez, G., Boyer, M., Fournous, G., and Raoult, D. (2011a). The giant *Cafeteria roenbergensis* virus that infects a widespread marine phagocytic protist is a new member of the fourth domain of life. *PLoS One* 6:e18935. doi: 10.1371/journal.pone.0018935

Colson, P., and Raoult, D. (2010). Gene repertoire of amoeba-associated giant viruses. *Intervirology* 53, 330–343. doi: 10.1159/000312918

Colson, P., Yutin, N., Shabalina, S. A., Robert, C., Fournous, G., La Scola, B., et al. (2011b). Viruses with more than 1,000 genes: Mamavirus, a new *Acanthamoeba polyphaga* mimivirus strain, and reannotation of mimivirus genes. *Genome Biol. Evol.* 3, 737–742. doi: 10.1093/gbe/evr048

Delaroque, N., and Boland, W. (2008). The genome of the brown alga *Ectocarpus siliculosus* contains a series of viral DNA pieces, suggesting an ancient association with large dsDNA viruses. *BMC Evol. Biol.* 8:320. doi: 10.1186/1471-2148-8-320

Derelle, E., Ferraz, C., Escande, M. L., Eychenié, S., Cooke, R., Piganeau, G., et al. (2008). Life-cycle and genome of OtV5, a Large DNA virus of the pelagic marine unicellular green alga Ostreococcus tauri. *PLoS One* 3:e02250. doi: 10.1371/journal.pone.0002250

Derelle, R., López-García, P., Timpano, H., and Moreira, D. (2016). A phylogenomic framework to study the diversity and evolution of stramenopiles (=Heterokonts). *Mol. Biol. Evol.* 33, 2890–2898. doi: 10.1093/molbev/msw168

Dong, S., Raffaele, S., and Kamoun, S. (2015). The two-speed genomes of filamentous pathogens: Waltz with plants. *Curr. Opin. Genet. Dev.* 35, 57–65. doi: 10.1016/j.gde.2015.09.001

Edgar, R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 32, 1792–1797.

Erwin, D. C., and Ribeiro, O. K. (1996). *Phytophthora* diseases worldwide. St Paul, Minnesota: APS Press.

Faure, C., Veyssière, M., Boëlle, B., Clemente, H. S., Bouchez, O., Lopez-Roques, C., et al. (2020). Long-read genome sequence of the sugar beet rhizosphere mycoparasite *Pythium oligandrum*. *G3 Genes Genomes Genet.* 10, 431–436. doi: 10.1534/g3.119.400746

Filée, J., Pouget, N., and Chandler, M. (2008). Phylogenetic evidence for extensive lateral acquisition of cellular genes by *Nucleocytoplasmic* large DNA viruses. *BMC Evol. Biol.* 8:320. doi: 10.1186/1471-2148-8-320

Fischer, M. G., and Suttle, C. A. (2011). A virophage at the origin of large DNA transposons. *Science* 332, 231–235.

Franceschetti, M., Maqbool, A., Jiménez-Dalmaroni, M. J., Pennington, H. G., Kamoun, S., and Banfield, M. J. (2017). Effectors of *Filamentous* Plant pathogens: commonalities amid diversity. *Microbiol. Mol. Biol. Rev.* 81, 1–17. doi: 10.1128/mmbr.00066-16

Frank, J. A., and Feschotte, C. (2017). Co-option of endogenous viral sequences for host cell function. *Curr. Opin. Virol.* 25, 81–89. doi: 10.1016/j.coviro.2017.07.021

Gallot-Lavallée, L., and Blanc, G. (2017). A glimpse of nucleo-cytoplasmic large DNA virus biodiversity through the eukaryotic genomicswindow. *Viruses* 9:17. doi: 10.3390/v9010017

Gaulin, E., Pel, M. J. C., Camborde, L., San-Clemente, H., Courbier, S., Dupouy, M. A., et al. (2018). Genomics analysis of *Aphanomyces* spp. identifies a new class of oomycete effector associated with host adaptation. *BMC Biol.* 16:43. doi: 10.1186/s12915-018-0508-5

Geering, A. D. W., Maumus, F., Copetti, D., Choisne, N., Zwickl, D. J., Zytnicki, M., et al. (2014). Endogenous florendoviruses are major components of plant genomes and hallmarks of virus evolution. *Nat. Commun.* 5, 1–11. doi: 10.1038/ncomms6269

Gilbert, C., Chateigner, A., Ernenwein, L., Barbe, V., Bézier, A., Herniou, E. A., et al. (2014). Population genomics supports baculoviruses as vectors of horizontal transfer of insect transposons. *Nat. Commun.* 5:3348. doi: 10.1038/ncomms4348

Haas, B. J., Kamoun, S., Zody, M. C., Jiang, R. H. Y., Handsaker, R. E., Cano, L. M., et al. (2009). Genome sequence and analysis of the Irish potato famine pathogen *Phytophthora infestans*. *Nature* 461, 393–398. doi: 10.1038/nature08358

Hingamp, P., Grimsley, N., Acinas, S. G., Clerissi, C., Subirana, L., Poulain, J., et al. (2013). Exploring nucleo-cytoplasmic large DNA viruses in Tara oceans microbial metagenomes. *ISME J.* 7, 1678–1695. doi: 10.1038/ismej.2013.59

Inoue, Y., Saga, T., Aikawa, T., Kumagai, M., Shimada, A., Kawaguchi, Y., et al. (2017). Complete fusion of a transposon and herpesvirus created the Teratorn mobile element in medaka fish. *Nat. Commun.* 8, 1–14. doi: 10.1038/s41467-017-00527-2

Iyer, L. M., Balaji, S., Koonin, E. V., and Aravind, L. (2006). Evolutionary genomics of nucleo-cytoplasmic large DNA viruses. *Virus Res.* 117, 156–184. doi: 10.1016/j.virusres.2006.01.009

Ji, Y., Thomas, C., Tulin, N., Lodhi, N., Boamah, E., Kolenko, V., et al. (2016). Charon mediates immune deficiency–driven PARP-1–dependent immune responses in *Drosophila*. *J. Immunol.* 197, 2382–2389. doi: 10.4049/jimmunol.1600994

Kamoun, S., Furzer, O., Jones, J. D. G., Judelson, H. S., Ali, G. S., Dalio, R. J. D., et al. (2015). The Top 10 oomycete pathogens in molecular plant pathology. *Mol. Plant Pathol.* 16, 413–434. doi: 10.1111/mpp.12190

Krupovic, M., Bamford, D. H., and Koonin, E. V. (2014). Conservation of major and minor jelly-roll capsid proteins in Polinton (Maverick) transposons suggests that they are bona fide viruses. *Biol. Direct* 9, 1–7. doi: 10.1186/1745-6150-9-6

Kumar, S., Stecher, G., Suleski, M., and Hedges, S. B. (2017). TimeTree: A resource for timelines, timetrees, and divergence times. *Mol. Biol. Evol.* 34, 1812–1819. doi: 10.1093/molbev/msx116

La Scola, B., Audic, S., Robert, C., Jungang, L., De Lamballerie, X., Drancourt, M., et al. (2003). A giant virus in amoebae. *Science* 299:2033. doi: 10.1126/science.1081867

La Scola, B., Desnues, C., Pagnier, I., Robert, C., Barrassi, L., Fournous, G., et al. (2008). The virophage as a unique parasite of the giant mimivirus. *Nature* 455, 100–104. doi: 10.1038/nature07218

Lacourt, I., Panabières, F., Marais, A., Venard, P., and Ricci, P. (1994). Intraspecific polymorphism of *Phytophthora parasitica* revealed by analysis of mitochondrial DNA restriction fragment length polymorphism. *Mycol. Res.* 98, 562–568. doi: 10.1016/S0953-7562(09)80480-2

Lang, D., Ullrich, K. K., Murat, F., Fuchs, J., Jenkins, J., Haas, F. B., et al. (2018). The *Physcomitrella patens* chromosome-scale assembly reveals moss genome structure and evolution. *Plant J.* 93, 515–533. doi: 10.1111/tpj.13801

Lara-Ramírez, R., Poncelet, G., Patthey, C., and Shimeld, S. M. (2017). The structure, splicing, synteny and expression of lamprey COE genes and the evolution of the COE gene family in chordates. *Dev. Genes Evol.* 227, 319–338. doi: 10.1007/s00427-017-0591-6

Laslett, D., and Canback, B. (2004). ARAGORN, a program to detect tRNA genes and tmRNA genes in nucleotide sequences. *Nucleic Acids Res.* 32, 11–16. doi: 10.1093/nar/gkh152

Leonard, G., Labarre, A., Milner, D. S., Monier, A., Soanes, D., Wideman, J. G., et al. (2018). Comparative genomic analysis of the 'pseudofungus' *Hyphochytrium catenoides*. *Open Biol.* 8:170184. doi: 10.1098/rsob.170184

Letunic, I., Khedkar, S., and Bork, P. (2020). SMART: recent updates, new developments and status in 2020. *Nucleic Acids Res.* 49, 458–460. doi: 10.1093/nar/gkaa937

Li, J., Mahajan, A., and Tsai, M. D. (2006). Ankyrin repeat: A unique motif mediating protein-protein interactions. *Biochemistry* 45, 15168–15178. doi: 10.1021/bi062188q

Liu, H., Ma, X., Yu, H., Fang, D., Li, Y., Wang, X., et al. (2016). Genomes and virulence difference between two physiological races of *Phytophthora nicotianae*. *Gigascience* 5, 1–8. doi: 10.1186/s13742-016-0108-7

Lowe, T. M., and Chan, P. P. (2016). tRNAscan-SE On-line: integrating search and context for analysis of transfer RNA genes. *Nucleic Acids Res.* 44, W54–W57. doi: 10.1093/nar/gkw413

Matari, N. H., and Blair, J. E. (2014). A multilocus timescale for oomycete evolution estimated under three distinct molecular clock models. *BMC Evol. Biol.* 14:101. doi: 10.1186/1471-2148-14-101

Maumus, F., Epert, A., Nogué, F., and Blanc, G. (2014). Plant genomes enclose footprints of past infections by giant virus relatives. *Nat. Commun.* 5, 1–10. doi: 10.1038/ncomms5268

Mccarthy, C. G. P., and Fitzpatrick, D. A. (2017). Phylogenomic reconstruction of the oomycete phylogeny derived from 37 genomes. *mSphere* 2, 1–17.

Mirzakhanyan, Y., and Gershon, P. D. (2020). Structure-based deep mining reveals first-time annotations for 46% of the dark annotation space of the 9,671-member super-proteome of the nucleo-cytoplasmic large DNA viruses. *J. Virol.* 94:e00854–20. doi: 10.1128/jvi.00854-20

Mistry, J., Chuguransky, S., Williams, L., Qureshi, M., Salazar, G. A., Sonnhammer, E. L. L., et al. (2020). Pfam: The protein families database in 2021. *Nucleic Acids Res.* 49, 412–419. doi: 10.1093/nar/gkaa913

Moniruzzaman, M., Martinez-Gutierrez, C. A., Weinheimer, A. R., and Aylward, F. O. (2020a). Dynamic genome evolution and complex virocell metabolism of globally-distributed giant viruses. *Nat. Commun.* 11, 1–11. doi: 10.1038/s41467-020-15507-2

Moniruzzaman, M., Weinheimer, A. R., Martinez-Gutierrez, C. A., and Aylward, F. O. (2020b). Widespread endogenization of giant viruses shapes genomes of green algae. *Nature* 588, 141–145. doi: 10.1038/s41586-020-2924-2

Mougari, S., Sahmi-Bounsiar, D., Levasseur, A., Colson, P., and Scola, B. L. (2019). Virophages of giant viruses: An update at eleven. *Viruses* 11, 1–28. doi: 10.3390/v11080733

Ngaki, M. N., Wang, B., Sahu, B. B., Srivastava, S. K., Farooqi, M. S., Kambakam, S., et al. (2016). Tanscriptomic study of the soybean-*Fusarium virguliforme* interaction revealed a novel ankyrin-repeat containing defense gene, expression of whose during infection led to enhanced resistance to the fungal pathogen in transgenic soybean plants. *PLoS One* 11:e163106. doi: 10.1371/journal.pone.0163106

Pagnier, I., Reteno, D. G. I., Saadi, H., Boughalmi, M., Gaia, M., Slimani, M., et al. (2013). A decade of improvements in mimiviridae and marseilleviridae isolation from amoeba. *Intervirology* 56, 354–363. doi: 10.1159/000354556

Panabières, F., Ali, G. S., Allagui, M. B., Dalio, R. J. D., Gudmestad, N. C., Kuhn, M.-L., et al. (2016). *Phytophthora nicotianae* diseases worldwide: new knowledge of a long-recognised pathogen. *Phytopathol. Mediterr.* 55, 20–40. doi: 10.14601/Phytopathol_Mediterr-16423

Panabières, F., and Le Berre, J. Y. (1999). A family of repeated DNA in the genome of the oomycete plant pathogen P*hytophthora cryptogea*. *Curr. Genet.* 36, 105–112. doi: 10.1007/s002940050479

Panabières, F., Rancurel, C., da Rocha, M., and Kuhn, M. L. (2020). Characterization of two satellite DNA families in the genome of the oomycete plant pathogen *Phytophthora parasitica*. *Front. Genet.* 11:557. doi: 10.3389/fgene.2020.00557

Pandurangan, A. P., Stahlhacke, J., Oates, M. E., Smithers, B., and Gough, J. (2019). The SUPERFAMILY 2.0 database: A significant proteome update and a new webserver. *Nucleic Acids Res.* 47, D490–D494. doi: 10.1093/nar/gky1130

Price, M. N., Dehal, P. S., and Arkin, A. P. (2010). FastTree 2 - Approximately maximum-likelihood trees for large alignments. *PLoS One* 5:e09490. doi: 10.1371/journal.pone.0009490

Pritham, E. J., Putliwala, T., and Feschotte, C. (2007). Mavericks, a novel class of giant transposable elements widespread in eukaryotes and related to DNA viruses. *Gene* 390, 3–17. doi: 10.1016/j.gene.2006.08.008

Raoult, D., Audic, S., Robert, C., Abergel, C., Renesto, P., Ogata, H., et al. (2004). The 1.2-Megabase Genome Sequence of Mimivirus. *Science* 306, 1344–1350.

Raoult, D., and Boyer, M. (2010). Amoebae as genitors and reservoirs of giant viruses. *Intervirology* 53, 321–329. doi: 10.1159/000312917

Rennoll-Bankert, K. E., Garcia-Garcia, J. C., Sinclair, S., and Dumler, J. S. (2015). Chromatin-bound bacterial effector AnkA recruits HDAC1 and modifies host gene expression. *Cell. Microbiol.* 17, 1640–1652. doi: 10.1111/cmi.12461

Reteno, D. G., Benamar, S., Khalil, J. B., Andreani, J., Armstrong, N., Klose, T., et al. (2015). Faustovirus, an Asfarvirus-related new lineage of giant viruses infecting Amoebae. *J. Virol.* 89, 6585–6594. doi: 10.1128/jvi.00115-15

Rozenberg, A., Oppermann, J., Wietek, J., Fernandez Lahore, R. G., Sandaa, R. A., Bratbak, G., et al. (2020). Lateral gene transfer of anion-conducting cannelrhodopsins between green algae and giant viruses. *Curr. Biol.* 30, 4910–4920. doi: 10.1016/j.cub.2020.09.056

Savory, F., Leonard, G., and Richards, T. A. (2015). The role of horizontal gene transfer in the evolution of the oomycetes. *PLoS Pathog.* 11:e1004805. doi: 10.1371/journal.ppat.1004805

Schulz, F., Alteio, L., Goudeau, D., Ryan, E. M., Yu, F. B., Malmstrom, R. R., et al. (2018). Hidden diversity of soil giant viruses. *Nat. Commun.* 9, 1–9. doi: 10.1038/s41467-018-07335-2

Schulz, F., Roux, S., Paez-Espino, D., Jungbluth, S., Walsh, D. A., Denef, V. J., et al. (2020). Giant virus diversity and host interactions through global metagenomics. *Nature* 578, 432–436. doi: 10.1038/s41586-020-1957-x

Schulz, F., Yutin, N., Ivanova, N. N., Ortega, D. R., Lee, T. K., Vierheilig, J., et al. (2017). Giant viruses with an expanded complement of translation system components. *Science* 356, 82–85.

Seidl, M. F., and Thomma, B. P. H. J. (2017). Transposable elements direct the coevolution between plants and microbes. *Trends Genet.* 33, 842–851. doi: 10.1016/j.tig.2017.07.003

Sharma, V., Colson, P., Giorgi, R., Pontarotti, P., and Raoult, D. (2014). DNA-dependent RNA polymerase detects hidden giant viruses in published databanks. *Genome Biol. Evol.* 6, 1603–1610. doi: 10.1093/gbe/evu128

Stam, R., Howden, A. J. M., Delgado-Cerezo, M., Amaro, T. M. M. M., Motion, G. B., Pham, J., et al. (2013). Characterization of cell death inducing *Phytophthora capsici* CRN effectors suggests diverse activities in the host nucleus. *Front. Plant Sci.* 4:387. doi: 10.3389/fpls.2013.00387

Uzuhashi, S., Kakishima, M., and Tojo, M. (2010). Phylogeny of the genus *Pythium* and description of new genera. *Mycoscience* 51, 337–365.

Vardi, A., Haramaty, L., Van Mooy, B. A. S., Fredricks, H. F., Kimmance, S. A., Larsen, A., et al. (2012). Host-virus dynamics and subcellular controls of cell fate in a natural coccolithophore population. *Proc. Natl. Acad. Sci. U. S. A.* 109, 19327–19332. doi: 10.1073/pnas.1208895109

Vetukuri, R. R., Åsman, A. K. M., Tellgren-Roth, C., Jahan, S. N., Reimegård, J., Fogelqvist, J., et al. (2012). Evidence for small RNAs homologous to effector-encoding genes and transposable elements in the oomycete *Phytophthora* infestans. *PLoS One* 7:e51399. doi: 10.1371/journal.pone.0051399

Wicker, P., Sabot, F., Hua-Van, A., Bennetzen, J. L., Capy, P., Chalhoub, B., et al. (2007). A unified classification system for eukaryotic transposable elements. *Nat. Genet.* 8, 973–982.

Yan, H.-Z., and Liou, R.-F. (2006). Selection of internal control genes for real-time quantitative RT-PCR assays in the oomycete plant pathogen *Phytophthora parasitica*. *Fungal Genet. Biol.* 6, 430–438. doi: 10.1016/j.fgb.2006.01.010

Yang, M., Duan, S., Mei, X., Huang, H., Chen, W., Liu, Y., et al. (2018). The *Phytophthora cactorum* genome provides insights into the adaptation to host defense compounds and fungicides. *Sci. Rep.* 8, 1–11. doi: 10.1038/s41598-018-24939-2

Yang, X., Tyler, B. M., and Hong, C. (2017). An expanded phylogeny for the genus *Phytophthora*. *IMA Fungus* 8, 355–384. doi: 10.5598/imafungus.2017.08.02.09

Yoshikawa, G., Blanc-Mathieu, R., Song, C., Kayama, Y., Mochizuki, T., Murata, K., et al. (2019). Medusavirus, a novel large DNA virus discovered from hot spring water. *J. Virol.* 93, 1–25.

Yutin, N., Wolf, Y. I., and Koonin, E. V. (2014). Origin of giant viruses from smaller DNA viruses not from a fourth domain of cellular life. *Virology* 466–467, 38–52. doi: 10.1016/j.virol.2014.06.032

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Microbiology

REVIEW
published: 02 June 2021
doi: 10.3389/fmicb.2021.648731

Marseilleviruses: An Update in 2021

Dehia Sahmi-Bounsiar[1,2†], Clara Rolland[1,2†], Sarah Aherfi[1,2], Hadjer Boudjemaa[1,3], Anthony Levasseur[1,2], Bernard La Scola[1,2] and Philippe Colson[1,2]*

[1] IHU Méditerranée Infection, Marseille, France, [2] Institut de Recherche pour le Développement (IRD), Assistance Publique-Hôpitaux de Marseille (AP-HM), MEPHI, Aix-Marseille Université, Marseille, France, [3] Department of Biology, Faculty of Natural Science and Life, Hassiba Benbouali University of Chlef, Chlef, Algeria

The family *Marseilleviridae* was the second family of giant viruses that was described in 2013, after the family *Mimiviridae*. *Marseillevirus marseillevirus*, isolated in 2007 by coculture on *Acanthamoeba polyphaga*, is the prototype member of this family. Afterward, the worldwide distribution of marseilleviruses was revealed through their isolation from samples of various types and sources. Thus, 62 were isolated from environmental water, one from soil, one from a dipteran, one from mussels, and two from asymptomatic humans, which led to the description of 67 marseillevirus isolates, including 21 by the IHU Méditerranée Infection in France. Recently, five marseillevirus genomes were assembled from deep sea sediment in Norway. Isolated marseilleviruses have ≈250 nm long icosahedral capsids and 348–404 kilobase long mosaic genomes that encode 386–545 predicted proteins. Comparative genomic analyses indicate that the family *Marseilleviridae* includes five lineages and possesses a pangenome composed of 3,082 clusters of genes. The detection of marseilleviruses in both symptomatic and asymptomatic humans in stool, blood, and lymph nodes, and an up-to-30-day persistence of marseillevirus in rats and mice, raise questions concerning their possible clinical significance that are still under investigation.

Keywords: marseillevirus, *Marseilleviridae*, giant virus, amoeba, *Pimascovirales*, megavirales, human

OPEN ACCESS

Edited by:
Masaharu Takemura,
Tokyo University of Science, Japan

Reviewed by:
Matthias Fischer,
Max-Planck-Gesellschaft (MPG), Germany
Hiroyuki Ogata,
Kyoto University, Japan

***Correspondence:**
Philippe Colson
philippe.colson@univ-amu.fr

† These authors have contributed equally to this work

Specialty section:
This article was submitted to Virology, a section of the journal Frontiers in Microbiology

Received: *01 January 2021*
Accepted: *12 April 2021*
Published: *02 June 2021*

Citation:
Sahmi-Bounsiar D, Rolland C, Aherfi S, Boudjemaa H, Levasseur A, La Scola B and Colson P (2021) Marseilleviruses: An Update in 2021. Front. Microbiol. 12:648731. doi: 10.3389/fmicb.2021.648731

INTRODUCTION

Defining any element is based on data and tools that are available at the moment, and definitions can evolve with technological progress (Popper, 2005). The first giant virus was isolated using culturing on *Acanthamoeba* sp. from a water sample collected in 1992 from a cooling tower in Bradford, United Kingdom and identified as a giant virus in 2003. It was initially presumed to be a bacteria and named *Bradford coccus* by T. J. Rowbotham, but it appeared to be a virus, with a particle size of ≈500 nm, named *Acanthamoeba polyphaga Mimivirus* (APMV) (La Scola et al., 2003). This virus was in total opposition to the concept of a virus, defined as a small particle invisible under light microscopy and ultrafilterable through pores with a diameter of 0.2 μm, and with a genetic armamentarium that usually did not exceed a few genes (Lwoff, 1957; La Scola et al., 2003; Raoult et al., 2007; Sharma et al., 2016). This discovery revolutionized the world of virology, sparked the curiosity of many scientists and launched an open debate about the definition and classification of viruses and the universal tree of Life (Raoult et al., 2004; Raoult and Forterre, 2008;

Moreira and López-García, 2009; Boyer et al., 2010; Forterre, 2010; Raoult, 2013). Giant virus evolution remains controversial among researchers. Giant viruses of amoeba other than Mimivirus were described (Aherfi et al., 2016b; Colson et al., 2017; Abergel and Claverie, 2020), and they were also related to a monophyletic group of viruses known as nucleocytoplasmic large DNA viruses (NCLDVs) that was described in 2001 and comprises poxviruses, asfaviruses, ascoviruses, iridoviruses, and phycodnaviruses, which have been the subjects of many studies before the discovery of APMV (Wardley et al., 1983; Iyer et al., 2001; Stasiak et al., 2003; Van Etten, 2003; Lefkowitz et al., 2006; Colson et al., 2012; Yutin and Koonin, 2012). We consider giant viruses as having a virion size > 200 nm (Sharma et al., 2016). In 2012, it was proposed to reclassify families of giant amoebal viruses in a new viral order, the Megavirales (Colson et al., 2013a). Since the end of 2019, the International Committee of Taxonomy of Viruses (ICTV) has officially classified giant viruses in the class *Megaviricetes* as part of the phylum *Nucleocytoviricota*, which is part of the kingdom *Bamfordvirae* in the realm *Varidnaviria* (Walker et al., 2019).

Acanthamoeba polyphaga Mimivirus is the first described giant virus and was serendipitously isolated using a coculture strategy on *Acanthamoeba polyphaga* (La Scola et al., 2003). It is the founder of a new viral family officially recognized by the ICTV and named *Mimiviridae* (La Scola et al., 2005a; Suzan-Monti et al., 2006). More than 100 mimivirus strains were thereafter isolated in *Acanthamoeba* spp., and classified in three lineages; A, B and C (Colson et al., 2012).

Marseillevirus (strain T19, later named *Marseillevirus marseillevirus*[1]) was discovered in 2007 as part of the continuation of research on giant amoeba viruses, by culturing on *Acanthamoeba polyphaga*. It is smaller, in terms of particle and genome sizes, than APMV (Boyer et al., 2009). It is the founder of a new viral family, officially recognized and named *Marseilleviridae* (Colson et al., 2013c), which has expanded over the last decade, with more than 50 members isolated; essentially from water, then insects, mussels, but also from humans. These samples were collected from seven countries, over the five continents: Europe, Africa, America, Oceania, and Asia. Marseilleviruses have been classified in the *Megaviricetes* (Walker et al., 2019). They represent a homogeneous group of giant viruses, although they can differ slightly from each other in their genomes and proteomes, the size of their particles, the morphology of the virions and their replicative cycle. After the discovery of mimiviruses, they were the first new giant viruses discovered and preceded numerous others (Colson et al., 2017); among them pandoraviruses (Philippe et al., 2013), pithoviruses (Legendre et al., 2014), faustoviruses (Reteno et al., 2015), molliviruses (Legendre et al., 2015), cedratviruses (Andreani et al., 2016), Pacmanvirus (Andreani et al., 2017), tupanviruses (Abrahão et al., 2018b), and Orpheovirus (Andreani et al., 2018). La Scola et al. also discovered a new type of virus, named virophage, able to parasitize mimiviruses by replicating in their viral factory and integrating in their genome (La Scola et al., 2008; Desnues and Raoult, 2012).

This review presents the most exciting discoveries and developments of the past 10 years concerning marseilleviruses.

MARSEILLEVIRUSES: DISCOVERY AND HISTORY

In 2007, Marseillevirus T19 (**Figure 1**) was isolated in Marseille by a coculture experiment on *Acanthamoeba polyphaga* from a water sample collected in a cooling tower in Paris (Boyer et al., 2009). Cryo-electron microscopy showed a capsid with a diameter of approximately 250 nm, with a shell thickness ≈10 nm separating it from the nucleocapsid by a space ≈5 nm. The marseillevirus surface has fibers of 12 nm in length with globular ends. In 2005, the investigation of a water sample collected in the Seine river around Paris had previously led to the identification of a small intracellular Gimenez-positive coccus (Thomas et al., 2008). In 2011, a second analysis of this sample in a coculture on *Acanthamoeba castellanii* allowed the description of the second giant virus of the family *Marseilleviridae*, named lausannevirus. It has an icosahedral virion with a 190–220 nm diameter without fibrils (Thomas et al., 2011). From there, marseillevirus research in environmental water samples was accelerated, leading from 2013 to 2015 to the isolation of cannes 8 virus from water of a cooling tower in Cannes, southeastern France (Aherfi et al., 2013), and of tunisvirus and saint-charles viruses taken from fresh water collected in fountains in Ariana, a suburb of Tunis, Tunisia and in Marseille, France, respectively (Pagnier et al., 2013; Aherfi et al., 2014). Melbournevirus came from a freshwater pond in Melbourne, Australia (Doutre et al., 2014) and was remarkably similar to marseillevirus and cannes 8 virus. Port-miou virus was isolated from the brackish water flowing from a submarine karstic spring in Port-Miou, Cassis, Marseille, France (Doutre et al., 2015). In 2016, water and soil samples taken from a bank of the Arakawa River, Japan led to the characterization of tokyovirus (Takemura, 2016), and samples of sewage collected in the Pampulha lagoon, Brazil led to the characterization of Brazilian marseillevirus (Dornas et al., 2016). In 2017, noumeavirus (Fabre et al., 2017) and kurlavirus (Chatterjee and Kondabagil, 2017) were isolated from a muddy sample of fresh water collected in a pond near Noumea airport, New Caledonia and a sewage water sample collected in Mumbai, India, respectively. In 2018, two marseillevirus-like viruses were isolated from soil samples collected in an aboriginal village (Serendah village) in Peninsular Malaysia (Tan et al., 2018) and from a fresh water pond bed soil in Shanghai (GenBank MG827395). More recently, in 2019, 15 new members of marseilleviruses, classified in three groups -hokutoviruses, kashiwazakiviruses, and kyotoviruses- were isolated from three water samples from Japan (Aoki et al., 2019). The same year, 15 new marseilleviruses were discovered in sewage, swamp, or wastewater from different locations in Algeria (**Supplementary Table 1**). Finally, 14 novel marseillevirus isolates have been recently reported from five Japanese aquatic sampling locations, as confirmed by molecular phylogenetic analyses of the major

[1]https://talk.ictvonline.org/taxonomy/p/taxonomy-history?taxnode_id=201903801

capsid protein (MCP); 11 strains belonged to lineage B, two to lineage A, and one to another lineage close to tokyovirus (Aoki et al., 2021). Therefore, marseilleviruses, like other giant viruses and amoeba, have been shown to be common in water and soil worldwide.

In 2013, the internal organs and digestive tract of a larva from a dipteran, *Eristalis tenax,* were screened. This larva was one of 86 larvae collected from two reservoirs of stagnant water in Tunis, Tunisia, and it allowed isolation of the first giant virus from an insect, which was named insectomime virus (Boughalmi et al., 2013). It was not detected on the surface of the larva, which could indicate the possibility of direct ingestion of the virus or an ingestion of an infected amoeba. Amoeba and non-giant viruses were previously described in insects (Evans and Schwarz, 2011; Otta et al., 2012; Li et al., 2014; Muli et al., 2014).

As marseilleviruses are common in water, golden mussels (Limnoperna fortunei) in southern Brazil were investigated after harvesting them in 2016 from a metal grill that had remained submerged at two meters for 6 months in Guaíba Lake. Decontamination of the surface of the mussels was done, the inner water was collected (dos Santos et al., 2016). Then a coculture procedure of the samples with *Acanthamoeba polyphaga* and *Acanthamoeba castellanii* was performed in order to identify the virus specificity for both species of amoeba. Transmission electron microscopy revealed an icosahedral particle of about 200 nm in diameter named golden marseillevirus (dos Santos et al., 2016). Non-giant viruses and amoeba were also previously described in mollusks, which could therefore be used as an indicator of biological contamination (Mosteo et al., 2016; Ilic et al., 2017).

Senegalvirus is the first giant virus isolated from a human, being revealed as one of the components of the gut microbiota. It was isolated from a stool sample which was collected in 2012 in N'Diop, a rural village in Senegal, from a 20-year-old asymptomatic Senegalese man (Lagier et al., 2012) using the coculture strategy on *Acanthamoeba* spp. (Lagier et al., 2012) and metagenomics (Colson et al., 2013b). Electron microscopy revealed the presence of icosahedral particles with a 196 nm diameter (Lagier et al., 2012). Giant blood marseillevirus is the second member of *Marseilleviridae* to be detected in humans, from the blood of a blood donor in 2010 in Marseille, France (Popgeorgiev et al., 2013a). In this study, complete genome sequencing, antigen detection, transmission electron microscopy, fluorescence *in situ* hybridization and cell culture were used and allowed giant blood marseillevirus detection in blood and in inoculated human lymphocytes (Popgeorgiev et al., 2013a). These discoveries suggest that the study of the human virome is still in its infancy (**Figure 2**).

Currently, metagenomics allows massive investigations of the virome and the microbiome, in addition to studies conducted by co-culture. Genomes from members of the family *Mimiviridae* and other giant virus relatives had already been discovered with this technique (Schulz et al., 2017, 2018). In 2019, for the first time, sequences from five marseillevirus-like viruses were retrieved in deep sea sediment near the Loki's Castle hydrothermal vent field in Norway (Bäckström et al., 2019).

REPLICATION CYCLE ANALYSIS OF MARSEILLEVIRUSES

Marseillevirus and Its Different Pathways for Entering Amoeba

Phagocytosis by *Acanthamoeba* is triggered by particles greater than 500 nm (Weisman and Korn, 1967). It is used by amoebae, and macrophages, to capture mimiviruses (La Scola et al., 2003; Ghigo et al., 2008). The fact that marseilleviruses have a diameter of 250 nm makes this process less likely to occur, although it was described during their discovery (Boyer et al., 2009). In 2016 a study on marseilleviruses and their method for penetrating amoeba was done and revealed different viral strategies for amoeba penetration (Arantes et al., 2016). Scanning and transmission electron microscopy was used to characterized giant vesicles, measuring from 300 nm to more than 1 μm in diameter, depending on the number of viral particles that were encompassed, from dozens to thousands. These vesicles have one or more membranes, likely originating from the endoplasmic reticulum. An interaction of *Acanthamoeba castellanii* with vesicles was reported after 30 min of incubation. When a vesicle has only one membrane, it fuses with the phagosome membrane and releases virus particles inside the amoebal cytoplasm. Alternatively, when vesicles have several membranes, only the external one is merged with the phagosome membrane, keeping vesicles intact inside the amoeba cytoplasm. Another possible way for entry into the amoeba cytoplasm is by phagocytosis of groups of marseillevirus particles, or the endocytosis of single particles (Arantes et al., 2016). **Figure 3** summarizes the different pathways of entry of marseilleviruses into amoeba and schematizes the marseillevirus replication cycle.

Cytopathogenic Effect of Marseilleviruses on Amoeba

During the period from 2007 to 2021 (Rolland et al., 2019), all marseilleviruses were reported to induce the same cytopathogenic effect (CPE) on amoeba; i.e., the rounding of the amoeba and its lysis. Among the 15 marseilleviruses most recently discovered in Japan (Aoki et al., 2019) and classified into 3 groups, two of them: kashiwazakivirus and hokutovirus, belonging to lineage B, presented, in addition to cell rounding, another CPE consisting of aggregation with uninfected cells, which promotes viral dissemination and induces "bunch" formation of amoeba and is original in the family *Marseilleviridae,* but was previously observed with tupanvirus, a mimivirus (Oliveira et al., 2019). Aoki et al. have recently studied the involvement of monosaccharides in bunch formation by amoeba cells infected with a marseillevirus, and they reported that this was observed in lineage B but not lineage A strains. In addition, they observed that galactose inhibited bunch formation whereas mannose and glucose did not.

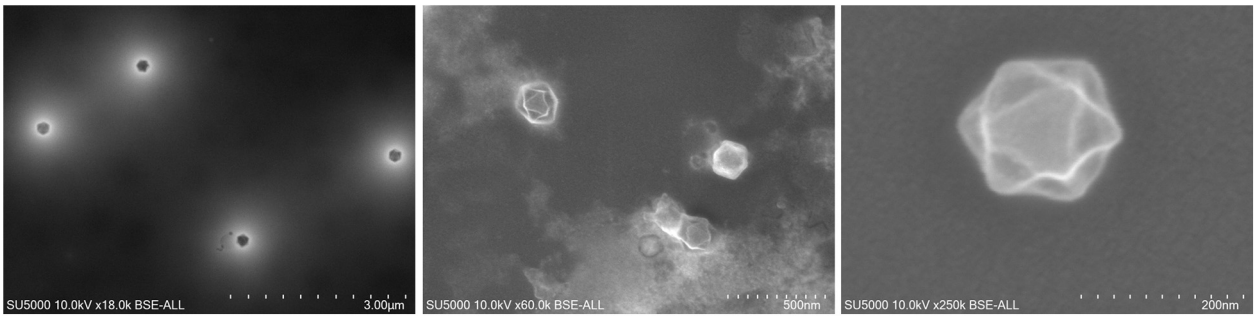


FIGURE 1 | Scanning electron microscopy of marseillevirus T19 particles. Micrographs of viral particles from culture supernatant. Technical settings and scale bars are indicated in the images.

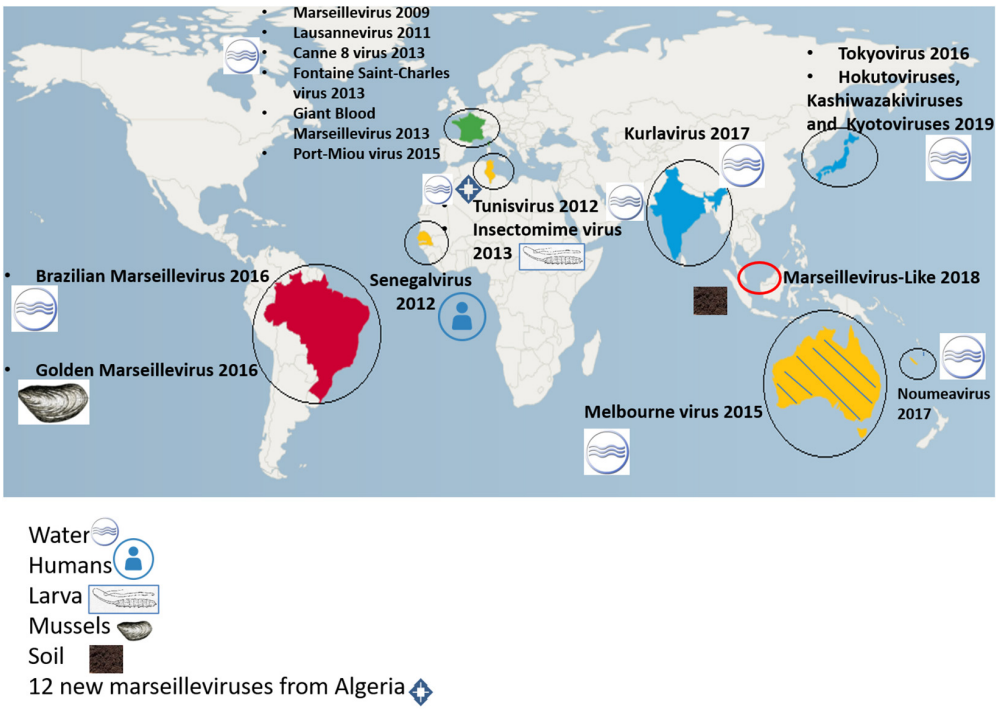


FIGURE 2 | Worldwide distribution of isolated marseilleviruses.

Morphogenesis and Release of Virions

All marseilleviruses have been reported to display characteristics for their replication and virion morphogenesis similar to *Marseillevirus marseillevirus*, although with some distinctive features. These steps occur in the cytoplasm in viral factories, which are diffuse areas near the amoeba nucleus. For marseillevirus, encapsidation of viral DNA was described as subsequent to the formation of mature and immature particles (Boyer et al., 2009). The replication cycle was completed within 12 h, with an eclipse phase 2 h post-infection (p.i.) followed by the appearance of a large viral factory in the host cytoplasm at 4 h p.i., by virion morphogenesis inside and around the factory, and by complete lysis of amoeba at 12 h p.i., which released new viral progeny able to infect other amoeba. Fabre et al. (2017) studied the replication of noumeavirus and reported that it initiates its replication by recruiting the transcription machinery of the amoebal nucleus to their viral factory located in the cytoplasm, while its virions did not incorporate the transcription machinery encoded by their own genes. In fact, it was reported that recruitment of nuclear proteins such as GFP-SUMO from the host nucleus occurred in the early times of the virus replication. These proteins could have essential functions such as the early transcription of mRNA as well as the polyadenylation of viral transcripts. These virus-induced changes, although transient, may allow the viral machinery to take over the host. Giant blood marseillevirus was observed 21 days after inoculation in Jurkat cells, an immortalized line of human T lymphocytes, but did not

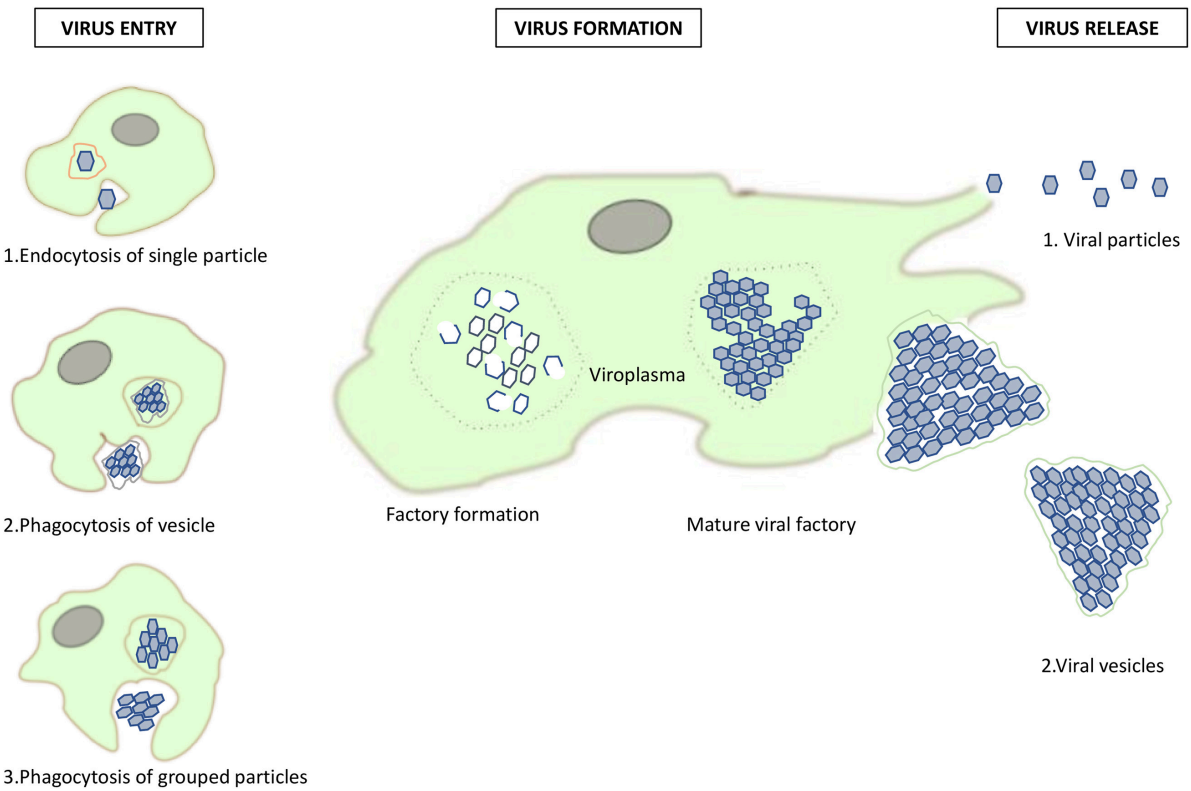


FIGURE 3 | Replication cycle of marseilleviruses. A summary of the different pathways of marseillevirus entry and new virion assembly in amoeba, and their release pathways.

propagate (Popgeorgiev et al., 2013a). This suggested a possible larger host spectrum for marseilleviruses, not only restricted to amoeba. No virophage, such as those previously described infecting mimiviruses of different lineages (La Scola et al., 2008), has been observed in association with marseilleviruses. It has been hypothesized that the fibrils of mimiviruses were essential for virophage entry (Desnues and Raoult, 2010); also the presence in mimiviruses of protein R135 was associated with fibers (Sobhy et al., 2015). The lack of long fibrils and protein R135 or homologs in marseilleviruses might explain the incapacity of virophages to infect them, but this needs to be further investigated.

GENOMIC STUDY OF MARSEILLEVIRUSES

General Genomic Characteristics

The genomes of marseilleviruses are double stranded DNA with a size ranging from 348 to 404 kilobases (kb) and with a G + C content between 42.9 and 44.8%. The number of predicted genes varies between 386 for kurlavirus and 491 for Brazilian marseillevirus. A high proportion of these genes corresponds to ORFans (genes without significantly similar sequences in databases) or hypothetical proteins with no function assigned. This is a common feature of giant viruses (Raoult and Boyer, 2010). Several paralogous families of proteins are described in marseillevirus genomes. The most abundant are MORN (Membrane Occupation and Recognition Nexus) repeat-containing proteins, serine and/or threonine kinases (a host-signaling protein family), restriction endonucleases, F-box containing proteins, ankyrin repeat containing proteins, and HNH endonucleases. Interestingly, some particularities were discovered in comparison with other giant viruses such as mimiviruses. Notably, marseilleviruses do not possess elements of translational apparatus. Moreover, three fused genes whose products were identified as homologs of histone-like proteins (doublet-histone H2B/H2A, unknown domain/H2A and archaeal histone/H3) were found in marseilleviruses, which might be involved in viral DNA compaction, protection and/or regulation (Thomas et al., 2011; Erives, 2017).

The high rate of mosaicism observed in *Marseilleviridae* genomes, even in comparison with that of other giant viruses, is singular and intriguing (Boyer et al., 2009). The sympatric life within amoeba for viruses, bacteria, and symbionts facilitates the transfer of genetic material between these microorganisms. This melting pot within amoeba most likely explains the diverse origins of the genes noticed in marseillevirus genomes. In fact, the genes have various putative origins and have been involved in putative exchanges with viruses, essentially giant ones, and bacteria, archaea and eukaryotes, including *Acanthamoeba* spp., the marseillevirus hosts.

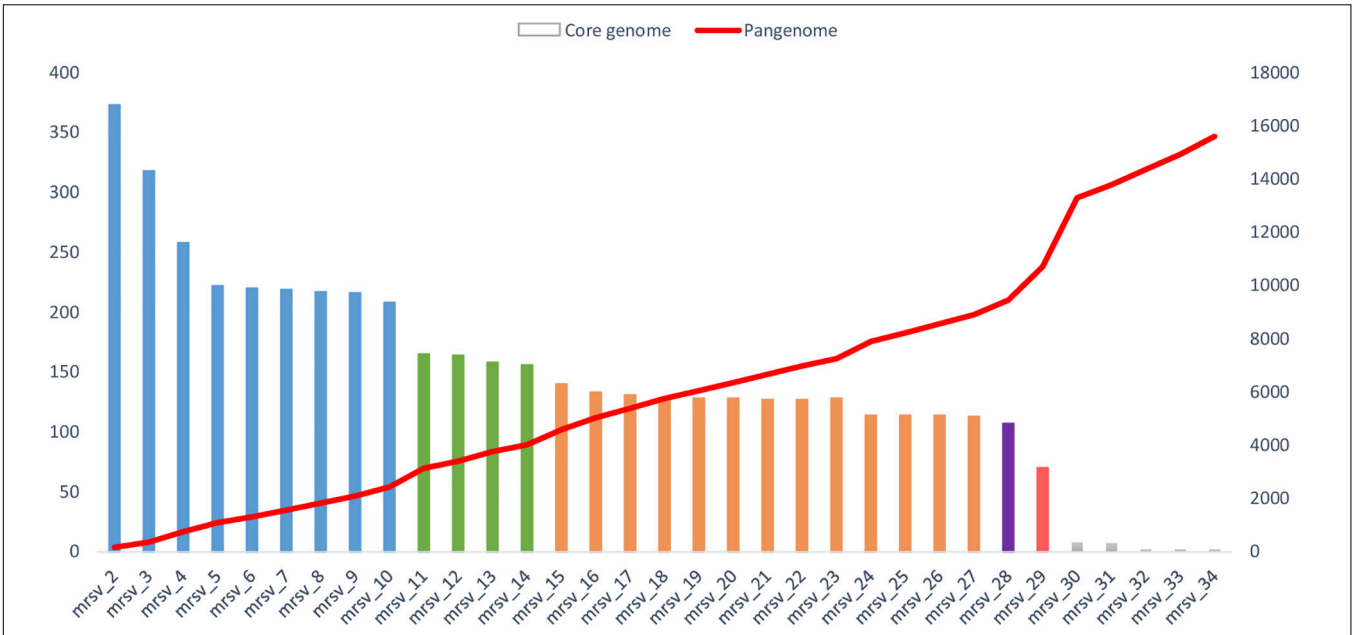


FIGURE 4 | Evolution of the size of the core genome and pangenome of marseilleviruses. The analysis was made by a reciprocal best BLAST hit strategy with significant thresholds for amino acid identity and coverage of 30 and 70%, respectively. The core genome is represented by the histogram column. Each lineage is labeled by a specific color: blue for lineage A, green for lineage B, orange for lineage C, purple for lineage D, and red for lineage E. The pangenome is highlighted by the red curve. The order for inclusion of each marseillevirus in the analysis corresponds to the order shown in **Supplementary Table 1**, as follows: mrsv_2: Cannes 8 virus; mrsv_3: Melbournevirus; mrsv_4: Tokyovirus A1; mrsv_5: Senegalvirus; mrsv_6: Phoenicianvirus; mrsv_7: Marseillevirus G648; mrsv_8: Marseillevirus G649; mrsv_9: Marseillevirus G650; mrsv_10: Marseillevirus Shangai 1; mrsv_11: Lausannevirus; mrsv_12: Noumeavirus; mrsv_13: Port-miou virus; mrsv_14: Kurlavirus BKC-1; mrsv_15: Tunisvirus; mrsv_16: Insectomime virus; mrsv_17: Marseillevirus N1; mrsv_18: Marseillevirus N16; mrsv_19: Marseillevirus N36; mrsv_20: Marseillevirus N40; mrsv_21: Marseillevirus N50; mrsv_22: Marseillevirus N57; mrsv_23: Marseillevirus N60A; mrsv_24: Marseillevirus N60B; mrsv_25: Marseillevirus NAQ2; mrsv_26: Marseillevirus AM2; mrsv_27: Marseillevirus AM21; mrsv_28: Brazilian marseillevirus; mrsv_29: Golden marseillevirus; mrsv_30: Marseillevirus LCMAC 101; mrsv_31: Marseillevirus LCMAC 102; mrsv_32: Marseillevirus LCMAC 103; mrsv_33: Marseillevirus LCMAC 201; mrsv_34: Marseillevirus LCMAC 202.

A recent study conducted on the noumeavirus DNA confirmed by pulse-field gel electrophoresis and restriction digestion that its genome is circular (Blanca et al., 2020). In addition, these authors conducted a comparative genomic study to analyze the organization of the marseillevirus genomes and reported an asymmetry in sequence conservation along these genomes. Two distinct genomic regions were detected including one that gathers most marseillevirus paralogous genes and underwent genomic rearrangements and another one that is enriched in core genes. In addition, a majority of the genes encoding proteins that compose the viral particles are located in the conserved genomic region. Moreover, the frequency of lately expressed genes was significantly greater in the core region (about two thirds of the genes) than in the other part of the genome (44% of the genes), indicating that genes of the core region are mostly expressed during the late phase of the replication cycle.

A study on promoters performed by Oliveira et al. (2017) showed the abundance of an AT-rich motif (AAATATTT) in multiple copies in intergenic regions of the marseillevirus genome. This motif was associated with 55% of marseillevirus genes, regardless of the lineage. Furthermore, variations in this putative promoter sequence were suspected to alter gene transcription. This investigation suggested that this motif might be an ancient characteristic of the family *Marseilleviridae* due to its abundance in the genomes, and that it might contribute to the genome mosaicism. However, no association of this motif with the temporal transcription levels of marseillevirus genes was observed in a transcriptomic study (Rodrigues et al., 2020).

Some transcripts were retrieved in the capsid of marseillevirus (Boyer et al., 2009). In addition, 49 proteins were identified by proteomics in marseillevirus virions, which comprised the capsid protein, the family B DNA polymerase and also the three histone-like proteins (Boyer et al., 2009). A recent analysis of the marseillevirus transcriptome identified a temporal transcription of genes (Rodrigues et al., 2020). Three categories of transcription profile were detected: early (0, 1, and 2 h p.i.), intermediate (4, 5, 6, and 8 h p.i.) and late expression (10 and 12 h p.i.). Furthermore, an alteration of the host transcription was also observed, with a global decrease in expression of amoeba genes, except for those related to exosome secretion, for which an increase was noted, suggesting the establishment of a defense mechanism by amoeba. It was suspected that the temporal transcription profile with a fast expression of transcription and translation factors was an adaptation of marseillevirus against the defense system developed by the infected amoeba. It is also worthy to note that DNA methylation was reported to be widespread among giant viruses, in two thirds of the giant viral families. Regarding marseilleviruses, some but not all were reported to encode complete sets of components of the restriction-modification (R-M) systems. As a matter of fact, all viruses belonging to lineage

A were found to encode a complete R-M system, but this was not the case for members of the other lineages (Jeudy et al., 2020).

Phylogeny

The family *Marseilleviridae* is classified in five lineages with, as the founding members, marseillevirus (Marseillevirus marseillevirus) in lineage A, and lausannevirus, tunisvirus, Brazilian marseillevirus, and golden marseillevirus in lineages B, C, D, and E, respectively (Boyer et al., 2009; Thomas et al., 2011; Aherfi et al., 2014; Dornas et al., 2016; dos Santos et al., 2016). A phylogenetic tree with most of the marseilleviruses available in the NCBI database was constructed (**Figure 5**). The MCP gene (the most widely-deposited one) was selected for tree construction with 49 marseilleviruses. The structure of the tree follows the previously-established classification, but the marseillevirus-like genomes generated from metagenomic data form two outgroups relative to the known lineages, one with marseillevirus LCMAC201 and marseillevirus LCMAC202, the

FIGURE 5 | Phylogenetic tree based on the major capsid protein (MCP). The analysis was performed using the Maximum Likelihood method (ML) with a JTT substitution matrix and 1,000 replicates. MCP sequences of marseilleviruses are color-labeled according to their classification into lineages: blue corresponds to lineage A, green to lineage B, orange to lineage C, purple to lineage D, and red to lineage E.

second with marseillevirus LCMAC 101, 102, and 103. Their positions between mamavirus used as an outgroup and all lineages of marseilleviruses indicate that they are evolutionarily distant from these latter. A majority of the new marseillevirus-like isolates from IHU Méditerranée Infection are clustered with tunisvirus and insectomime virus in lineage C, whereas the others, including phoenicianvirus, marseillevirus G648, G649, and G650, are clustered with marseilleviruses of lineage A. The 15 marseilleviruses discovered in Japan in 2019 were not yet officially classified. Congruent with Aoki et al. (2019) the seven kyoto virus members are grouped together in lineage A and the two hokutovirus members are grouped in lineage B. Regarding the six kashiwazakivirus members, five are positioned just outside lineage B but are closely related to this lineage, and the fifth one is part of lineage B. These results are congruent with the previous classification of kashiwazakiviruses as a subgroup of lineage B.

Pangenome Description

In this review, we updated the core and pangenome of the family *Marseilleviridae*. A total of 34 viruses were included in the analysis, with 15 newly-isolated marseilleviruses discovered at IHU Mediterranée Infection, including one already described in the literature (Schulz et al., 2020), five marseillevirus-like genomes built from metagenomic data (Bäckström et al., 2019), and 14 viruses already characterized belonging to the five *Marseilleviridae* lineages (**Supplementary Table 1**) (Boyer et al., 2009; Thomas et al., 2011; Aherfi et al., 2013, 2014; Boughalmi et al., 2013; Doutre et al., 2014, 2015; Dornas et al., 2016; dos Santos et al., 2016; Takemura, 2016; Chatterjee and Kondabagil, 2017; Fabre et al., 2017). The study was performed by the addition of the genomes one by one, lineage after lineage. First, we compared the genomes of the marseilleviruses from lineage A, then we added those from lineages B to E until the final inclusion of the five metagenome-derived marseillevirus genomes. The results showed an important difference with or without these latter marseillevirus genomes generated from metagenomic data. We present here the results in two parts, without and with the metagenome-derived genomes.

The total numbers of clusters of genes were 617, 499, and 512 for marseillevirus lineages A, B and C, respectively, and 75, 86, and 88% of these clusters included at least 2 genes for marseillevirus lineages A, B, and C, respectively (**Table 1** and **Figure 4**). All isolated marseillevirus genomes from all known lineages constituted a pangenome encompassing 1,011 clusters of genes, with 29% of them containing one sequence and 71% multiple sequences as determined using the CD-hit software, with thresholds of 30 and 50% for amino acid identity and coverage, respectively (Li and Godzik, 2006). The "strict" core genome (containing at least one sequence of each virus) included 121 clusters corresponding to 4,437 predicted proteins (**Supplementary Table 2**). It represented 12% of the pangenome. Different classes of proteins were identified in this core genome. The most represented is that of restriction endonucleases and of various other nucleases (including HNH homing endonucleases), with numerous paralogs in each virus. MORN repeat-containing proteins, ankyrin repeat proteins, and serine/threonine kinases also constituted a large number of these proteins. The marseillevirus core genes that are part of the giant virus core genome (Koonin and Yutin, 2010) consisted in 24 clusters of proteins and included the major capsid protein (MCP), the DNA topoisomerase II, the A32-like packaging ATPase, the D5 family helicase-primase, and DNA-directed RNA polymerase subunits. In addition, the clusters comprising the core genome included proteins annotated as putative glycosyltransferases, mannosyltransferase, metallopeptidase, RNA methyltransferase, and ATPase (**Supplementary Table 2**). The reciprocal best hits analysis performed with the ProteinOrtho software (v6.0.10) with amino acid identity and sequence coverage thresholds of 30 and 70%, respectively (Lechner et al., 2011) reflected the same increasing dynamic of pangenome size, and a total of 70 clusters of proteins included at least one sequence of each of the 29 viruses (0.6% of the full set).

The analysis of the pangenome reinforced the family *Marseilleviridae* classification by lineages. The pangenome size was relatively constant in each lineage (**Figure 3**). The addition of each lineage one by one coherently increased the pangenome size. Two points of substantial increase appeared between lineage A and B and between lineage A/B/C/D and lineage E. Nevertheless, this latter increase should be nuanced, due to the difference in the number of golden marseillevirus genes deposited in the NCBI database and the number described in the published article [296 versus 483, respectively (dos Santos et al., 2016)].

Metagenomes

The discovery of marseillevirus sequences in metagenomic data increased the genome size to 763 kb and the range of G + C content extended from 34.2 to 62.3% (**Supplementary Table 1**). Compared with the other members of the family *Marseilleviridae*, the number of predicted genes in genomes assembled from metagenomes ranged between 427 and a maximum of 793, which is unusual for marseilleviruses, for which the mean number of genes in isolated members is 438. The general picture of the gene composition of genomes assembled from metagenomes is similar to that of genomes from isolated marseilleviruses, with high proportions of ORFans, hypothetical proteins and other previously-described paralogous protein families, as well as a substantial level of mosaicism. In contrast, the presence of aminoacyl-tRNA synthetases and tRNAs were reported in metagenome-derived marseillevirus genomes, whereas they are absent in the genomes of isolated viruses (Bäckström et al., 2019). A total of 15 aminoacyl-tRNA synthetases and 26 tRNA were identified in the five metagenome-derived genomes. This is a brand-new perspective for marseilleviruses that merits confirmation in isolated strains.

The addition of the metagenome-derived marseillevirus genomes in the pangenome analysis dramatically modified the results. The pangenome size indeed reached 3,082 clusters of proteins. Among them, 61% contain a single predicted gene (**Table 1**). The "strict" core genome was drastically reduced compared to that based on genomes from isolated viruses, with 3 clusters of genes corresponding to the major capsid protein, the A32-like packaging ATPase and the ribonucleoside diphosphate

TABLE 1 | Pangenome and core genome size descriptions for marseilleviruses of lineage A to C and all viruses combined.

	Lineage A	Lineage B	Lineage C	All lineages	Metagenomes	All 34 marseilleviruses
Numbers of clusters (total)	617	499	512	1,011	2,086	3,082
Numbers of clusters with at least two genes	464 (75%)	428 (86%)	453 (88%)	714 (71%)	503 (24%)	1,199 (39%)
Numbers of clusters with one predicted gene	153 (25%)	71 (14%)	59 (12%)	297 (29%)	1,583 (76%)	1,883 (61%)
Strict Core genome	272	319	373	121	3	3

reductase small chain. The extreme reduction of the core genome was also observed by reciprocal best hits analysis (**Figure 3**).

MARSEILLEVIRUSES IN HUMANS

Marseilleviruses in Asymptomatic Humans

The giant viruses were associated with human pathology with the isolation of a mimivirus from two pneumonia patients, and experimental studies that reported that APMV could cause pneumonia (La Scola et al., 2005b; Saadi et al., 2013; Colson et al., 2016). In addition, the replication of APMV was reported in human macrophages and blood mononuclear cells (Ghigo et al., 2008; Silva et al., 2014), albeit this was not reproducible in the latter cells (Abrahão et al., 2018a) and no viral propagation was reported in these studies. Regarding marseilleviruses, a study of the microbial composition of the gut microbiota found, in the stool of an asymptomatic young Senegalese man, the first marseillevirus, which was named senegalvirus (Lagier et al., 2012). The giant blood marseillevirus was the second to be detected, by metagenomics, in the blood of an asymptomatic blood donor. In addition, PCR, fluorescence *in situ* hybridization (FISH) and electron microscopic analyses performed on Jurkat cells (T lymphocytes) indicated its presence (Popgeorgiev et al., 2013a). A seroprevalence study revealed IgG antibodies to marseillevirus in 13% of 174 blood donors from Marseille and Montpellier, France (Popgeorgiev et al., 2013b). In addition, IgG prevalence for marseillevirus was 23% in 22 thalassemia patients who repeatedly received blood transfusions, and marseillevirus DNA was detected in 9% of these thalassemia patients (Popgeorgiev et al., 2013b). In another study performed on 517 asymptomatic young adults sampled in Lausanne, Switzerland, the seroprevalence for lausannevirus examined by micro-immunofluorescence was 2.5% (Mueller et al., 2013). These results suggest that humans are frequently in contact with marseilleviruses. Moreover, an up-to-30-day persistence of marseillevirus in rats or mice was observed after inoculation through different inoculation routes: intraperitoneal, parenteral, or by aerosolization, and in deep organs (including liver, spleen, lung and nasal-associated lymphoid tissue); no evidence of disease was observed (Aherfi et al., 2018). However, a study of Italian patients showed no evidence of marseillevirus DNA in all 575 blood samples (from 285 healthy donors and 197 immunosuppressed patients) (Macera et al., 2019). Another study also reported the absence of marseillevirus DNA in blood donors and multi-transfused patients (Sauvage et al., 2014). Sequences identified as most related to marseillevirus were detected in metagenomes from asymptomatic people (Rampelli et al., 2016; Bzhalava et al., 2018).

Marseilleviruses in Symptomatic Humans

Marseilleviruses have also been observed in symptomatic humans. Antigens and DNA of marseillevirus were detected by immunohistochemistry and FISH analyses, respectively, in the lymph node of a febrile 11-month-old child with adenitis of unknown etiology (Popgeorgiev et al., 2013c). This study suggested that marseillevirus infection might become symptomatic in case of an immature or defective immune response causing adenitis. Marseillevirus DNA was also detected in pharyngeal and blood samples from a 20-year-old man with febrile gastroenteritis (Aherfi et al., 2016c). One year later, the patient was still positive to marseillevirus DNA. Detection of marseillevirus with PCR, FISH, direct immunofluorescence, and immunohistochemistry in the lymph node of a 30-year-old woman diagnosed with Hodgkin's lymphoma has also been reported in association with IgG antibodies to marseillevirus (Aherfi et al., 2016a). The causes of lymph node cancers are not fully understood; links with viruses and bacteria have been reported, with several possible mechanisms of lymphomagenesis such as for Epstein-Barr virus, *Helicobacter pylori,* and *Coxiella burnetii,* which is considered a factor that may promote lymphoid cancer. Previous data, indicating that blood transfusion could be a risk factor for the occurrence of lymphoma, question the possible involvement of marseillevirus in the occurrence of this disease. Therefore, these studies suggest that consideration should be given to marseilleviruses as additional viral causes of human pathology, but other investigations, including those by other teams, should work on this issue to improve our knowledge of the marseillevirus-human relationship. Interestingly, kinases from marseilleviruses were grouped with serine/threonine-protein kinase transforming proteins, and together with kinases from other giant viruses, were reported to display close similarity to viral oncogenes regarding the functional regions and putative substrate binding regions, questioning about a putative role in cancer (Janaki et al., 2020).

CONCLUSION

The family *Marseilleviridae* emerged in 2007 and expanded during the past decade to reach more than 50 members in 2021, being the second largest giant viral family after family *Mimiviridae*. During a decade marseilleviruses have been with mimiviruses the only known groups of giant viruses of

amoeba. Marseilleviruses have largely contributed, together with mimiviruses and the other giant viruses thereafter discovered, to delineate the remarkable features and define the criteria of what giant viruses are, and to point out how they differ from classical viruses. Although they are among the smallest of these giant viruses, they notwithstanding exhibit several specific phenotypic and genotypic features among giant viruses that warrant particular interest. One of them is for instance the presence of core histone-like genes that form sister clades to all four eukaryotic lineages of core histones. Marseilleviruses have a large worldwide distribution (nine countries, five continents), being found in samples such as water, soil, mussels, insects, and humans. Accordingly, new marseillevirus isolates continue to be described worldwide, including possible members from new lineages (Aoki et al., 2021). Their presence in both symptomatic and asymptomatic patients deserves further investigation.

AUTHOR CONTRIBUTIONS

PC, BL, and AL designed and supervised the review. DS-B and CR performed the review of the literature, collected the data, and performed the analyses. HB provided material. DS-B, CR, SA, AL, BL, and PC analyzed the data. DS-B, CR, and PC wrote the manuscript. All authors read the manuscript and approved the final version of the manuscript.

FUNDING

This work was supported by the French Government under the "Investments for the Future" program managed by the National Agency for Research (ANR), Méditerranée-Infection 10-IAHU-03. DS-B was financially supported through a grant from the Infectiopole Sud Foundation.

ACKNOWLEDGMENTS

This manuscript has been edited by a native English speaker.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2021.648731/full#supplementary-material

REFERENCES

Abergel, C., and Claverie, J. M. (2020). Giant viruses. *Curr. Biol.* 30, R1108–R1110. doi: 10.1016/j.cub.2020.08.055

Abrahão, J., Silva, L., Oliveira, D., and Almeida, G. (2018a). Lack of evidence of mimivirus replication in human PBMCs. *Microbes Infect.* 20, 281–283. doi: 10.1016/j.micinf.2018.03.003

Abrahão, J., Silva, L., Silva, L. S., Khalil, J. Y. B., Rodrigues, R., Arantes, T., et al. (2018b). Tailed giant *Tupanvirus* possesses the most complete translational apparatus of the known virosphere. *Nat. Commun.* 9:749. doi: 10.1038/s41467-018-03168-3161

Aherfi, S., Boughalmi, M., Pagnier, I., Fournous, G., La Scola, B., Raoult, D., et al. (2014). Complete genome sequence of *Tunisvirus*, a new member of the proposed family *Marseilleviridae*. *Arch. Virol.* 159, 2349–2358. doi: 10.1007/s00705-014-2023-2025

Aherfi, S., Colson, P., Audoly, G., Nappez, C., Xerri, L., Valensi, A., et al. (2016a). Marseillevirus in lymphoma: a giant in the lymph node. *Lancet Infect. Dis.* 16, e225–e234. doi: 10.1016/S1473-3099(16)30051-30052

Aherfi, S., Colson, P., La Scola, B., and Raoult, D. (2016b). Giant viruses of amoebas: an update. *Front. Microbiol.* 7:349. doi: 10.3389/fmicb.2016.00349

Aherfi, S., Colson, P., and Raoult, D. (2016c). *Marseillevirus* in the pharynx of a patient with neurologic disorders. *Emerg. Infect. Dis.* 22, 2008–2010. doi: 10.3201/eid2211.160189

Aherfi, S., Nappez, C., Lepidi, H., Bedotto, M., Barassi, L., Jardot, P., et al. (2018). Experimental inoculation in rats and mice by the giant *Marseillevirus* leads to long-term detection of virus. *Front. Microbiol.* 9:463. doi: 10.3389/fmicb.2018.00463

Aherfi, S., Pagnier, I., Fournous, G., Raoult, D., La Scola, B., and Colson, P. (2013). Complete genome sequence of Cannes 8 virus, a new member of the proposed family "*Marseilleviridae*.". *Virus Genes* 47, 550–555. doi: 10.1007/s11262-013-0965-964

Andreani, J., Aherfi, S., Khalil, J. Y. B., Di Pinto, F., Bitam, I., Raoult, D., et al. (2016). Cedratvirus, a double-cork structured giant virus, is a distant relative of pithoviruses. *Viruses* 8, 1–11. doi: 10.3390/v8110300

Andreani, J., Khalil, J. Y. B., Baptiste, E., Hasni, I., Michelle, C., Raoult, D., et al. (2018). Orpheovirus IHUMI-LCC2: a new virus among the giant viruses. *Front. Microbiol.* 8:2643. doi: 10.3389/fmicb.2017.02643

Andreani, J., Khalil, J. Y. B., Sevvana, M., Benamar, S., Di Pinto, F., Bitam, I., et al. (2017). Pacmanvirus, a new giant icosahedral virus at the crossroads between *Asfarviridae* and *Faustoviruses*. *J. Virol.* 91, e212–e217. doi: 10.1128/JVI.00212-217

Aoki, K., Fukaya, S., Takahashi, H., Kobayashi, M., Sasaki, K., and Takemura, M. (2021). *Marseilleviridae* lineage B diversity and bunch formation inhibited by galactose. *Microbes Environ.* 36:ME20139. doi: 10.1264/jsme2.ME20139

Aoki, K., Hagiwara, R., Akashi, M., Sasaki, K., Murata, K., Ogata, H., et al. (2019). Fifteen marseilleviruses newly isolated from three water samples in Japan reveal local diversity of *Marseilleviridae*. *Front. Microbiol.* 10:1152. doi: 10.3389/fmicb.2019.01152

Arantes, T. S., Rodrigues, R. A. L., Dos Santos, Silva, L. K., Oliveira, G. P., de Souza, H. L., et al. (2016). The large *Marseillevirus* explores different entry pathways by forming giant infectious vesicles. *J. Virol.* 90, 5246–5255. doi: 10.1128/JVI.00177-116

Bäckström, D., Yutin, N., Jørgensen, S. L., Dharamshi, J., Homa, F., Zaremba-Niedwiedzka, K., et al. (2019). Virus genomes from deep sea sediments expand the ocean megavirome and support independent origins of viral gigantism. *mBio* 10:e02497-18. doi: 10.1128/mBio.02497-2418

Blanca, L., Christo-Foroux, E., Rigou, S., and Legendre, M. (2020). Comparative analysis of the circular and highly asymmetrical *Marseilleviridae* genomes. *Viruses* 12:1270. doi: 10.3390/v12111270

Boughalmi, M., Pagnier, I., Aherfi, S., Colson, P., Raoult, D., and La Scola, B. (2013). First isolation of a marseillevirus in the diptera syrphidae eristalis tenax. *Intervirology* 56, 386–394. doi: 10.1159/000354560

Boyer, M., Yutin, N., Pagnier, I., Barrassi, L., Fournous, G., Espinosa, L., et al. (2009). Giant *Marseillevirus* highlights the role of amoebae as a melting pot in emergence of chimeric microorganisms. *Proc. Natl. Acad. Sci. U S A.* 106:21848. doi: 10.1073/pnas.0911354106

Boyer, M. L., Madoui, M.-A., Gimenez, G., La Scola, B., Raoult, D., and Martin, D. P. (2010). Phylogenetic and phyletic studies of informational genes in genomes highlight existence of a 4 th domain of life including giant viruses. *PLoS One* 5:e15530. doi: 10.1371/journal.pone.0015530

Bzhalava, Z., Hultin, E., and Dillner, J. (2018). Extension of the viral ecology in humans using viral profile hidden Markov models. *PLoS One* 13:e0190938. doi: 10.1371/journal.pone.0190938

Chatterjee, A., and Kondabagil, K. (2017). Complete genome sequence of Kurlavirus, a novel member of the family *Marseilleviridae* isolated in Mumbai. *India. Arch. Virol.* 162, 3243–3245. doi: 10.1007/s00705-017-3469-z

Colson, P., Aherfi, S., La Scola, B., and Raoult, D. (2016). The role of giant viruses of amoebas in humans. *Curr. Opin. Microbiol.* 31, 199–208. doi: 10.1016/j.mib.2016.04.012

Colson, P., de Lamballerie, X., Fournous, G., and Raoult, D. (2012). Reclassification of giant viruses composing a fourth domain of life in the new order *Megavirales*. *Intervirology* 55, 321–332. doi: 10.1159/000336562

Colson, P., de Lamballerie, X., Yutin, N., Asgari, S., Bigot, Y., Bideshi, D. K., et al. (2013a). "*Megavirales*", a proposed new order for eukaryotic nucleocytoplasmic large DNA viruses. *Arch. Virol.* 158, 2517–2521. doi: 10.1007/s00705-013-1768-1766

Colson, P., Fancello, L., Gimenez, G., Armougom, F., Desnues, C., Fournous, G., et al. (2013b). Evidence of the megavirome in humans. *J. Clin. Virol.* 57, 191–200. doi: 10.1016/j.jcv.2013.03.018

Colson, P., Pagnier, I., Yoosuf, N., Fournous, G., La Scola, B., and Raoult, D. (2013c). "*Marseilleviridae*", a new family of giant viruses infecting amoebae. *Arch. Virol.* 158, 915–920. doi: 10.1007/s00705-012-1537-y

Colson, P., La Scola, B., and Raoult, D. (2017). Giant viruses of amoebae: a journey through innovative research and paradigm changes. *Annu. Rev. Virol.* 4, 61–85. doi: 10.1146/annurev-virology-101416-141816

Desnues, C., and Raoult, D. (2010). Inside the lifestyle of the virophage. *Intervirology* 53, 293–303. doi: 10.1159/000312914

Desnues, C., and Raoult, D. (2012). Virophages question the existence of satellites. *Nat. Rev. Microbiol.* 10:234. doi: 10.1038/nrmicro2676-c3

Dornas, F. P., Assis, F. L., Aherfi, S., Arantes, T., Abrahão, J. S., Colson, P., et al. (2016). A Brazilian *Marseillevirus* is the founding member of a lineage in family *Marseilleviridae*. *Viruses* 8:76. doi: 10.3390/v8030076

dos Santos, R. N., Campos, F. S., Medeiros, de Albuquerque, N. R., Finoketti, F., Côrrea, R. A., et al. (2016). A new marseillevirus isolated in Southern Brazil from *Limnoperna fortunei*. *Sci. Rep.* 6:35237. doi: 10.1038/srep35237

Doutre, G., Arfib, B., Rochette, P., Claverie, J.-M., Bonin, P., and Abergel, C. (2015). Complete genome sequence of a new member of the *Marseilleviridae* recovered from the brackish submarine spring in the cassis port-miou calanque. *France. Genome Announc.* 3:e01148-15. doi: 10.1128/genomeA.01148-1115

Doutre, G., Philippe, N., Abergel, C., and Claverie, J.-M. (2014). Genome analysis of the first *Marseilleviridae* representative from Australia indicates that most of its genes contribute to virus fitness. *J. Virol.* 88, 14340–14349. doi: 10.1128/JVI.02414-2414

Erives, A. J. (2017). Phylogenetic analysis of the core histone doublet and DNA topo II genes of *Marseilleviridae*: evidence of proto-eukaryotic provenance. *Epigenetics Chromatin.* 10:55. doi: 10.1186/s13072-017-0162-160

Evans, J. D., and Schwarz, R. S. (2011). Bees brought to their knees: microbes affecting honey bee health. *Trends Microbiol.* 19, 614–620. doi: 10.1016/J.TIM.2011.09.003

Fabre, E., Jeudy, S., Santini, S., Legendre, M., Trauchessec, M., Couté, Y., et al. (2017). Noumeavirus replication relies on a transient remote control of the host nucleus. *Nat. Commun.* 8:15087. doi: 10.1038/ncomms15087

Forterre, P. (2010). Giant viruses: conflicts in revisiting the virus concept. *Intervirology* 53, 362–378. doi: 10.1159/000312921

Ghigo, E., Rgen Kartenbeck, J., Lien, P., Pelkmans, L., Capo, C., Mege, J.-L., et al. (2008). Ameobal pathogen mimivirus infects macrophages through phagocytosis. *PLoS Pathog* 4:e1000087. doi: 10.1371/journal.ppat.1000087

Ilic, N., Velebit, B., Teodorovic, V., Djordjevic, V., Karabasil, N., Vasilev, D., et al. (2017). Influence of environmental conditions on norovirus presence in mussels harvested in montenegro. *Food Environ. Virol.* 9, 406–414. doi: 10.1007/s12560-017-9298-9290

Iyer, L. M., Aravind, L., and Koonin, E. V. (2001). Common origin of four diverse families of large eukaryotic DNA viruses. *J. Virol.* 75, 11720–11734. doi: 10.1128/JVI.75.23.11720-11734.2001

Janaki, C., Malini, M., Tyagi, N., and Srinivasan, N. (2020). Unity and diversity among viral kinases. *Gene* 723:144134. doi: 10.1016/j.gene.2019.144134

Jeudy, S., Rigou, S., Alempic, J. M., Claverie, J. M., Abergel, C., and Legendre, M. (2020). The DNA methylation landscape of giant viruses. *Nat. Commun.* 2020:2657. doi: 10.1038/s41467-020-16414-2

Koonin, E. V., and Yutin, N. (2010). Origin and evolution of eukaryotic large nucleo-cytoplasmic DNA viruses. *Intervirology* 53, 284–292. doi: 10.1159/000312913

La Scola, B., Audic, S., Robert, C., Jungang, L., de Lamballerie, X., Drancourt, M., et al. (2003). A giant virus in amoebae. *Science* 299:2033. doi: 10.1126/science.1081867

La Scola, B., de Lamballerie, X. N., Claverie, J. M., Drancourt, M., and Raoult, D. (2005a). "Genus *Mimivirus*," in *Virus Taxonomy*, eds M. Fauquet, M. A. Mayo, J. Maniloff, U. Desselberger, and L. A. Ball (San Diego: Elsevier Academic Press), 275–276.

La Scola, B., Marrie, T. J., Auffray, J.-P., and Raoult, D. (2005b). Mimivirus in pneumonia patients. *Emerg. Infect. Dis.* 11, 449–452. doi: 10.3201/eid1103.040538

La Scola, B., Desnues, C., Pagnier, I., Robert, C., Barrassi, L., Fournous, G., et al. (2008). The virophage as a unique parasite of the giant mimivirus. *Nature* 455:100.

Lagier, J.-C., Armougom, F., Million, M., Hugon, P., Pagnier, I., Robert, C., et al. (2012). Microbial culturomics: paradigm shift in the human gut microbiome study. *Clin. Microbiol. Infect.* 18, 1185–1193. doi: 10.1111/1469-0691.12023

Lechner, M., Findeiß, S., Steiner, L., Marz, M., Stadler, P. F., and Prohaska, S. J. (2011). Proteinortho: detection of (Co-)orthologs in large-scale analysis. *BMC Bioinformatics* 12:124. doi: 10.1186/1471-2105-12-124

Lefkowitz, E. J., Wang, C., and Upton, C. (2006). Poxviruses: past, present and future. *Virus Res.* 117, 105–118. doi: 10.1016/j.virusres.2006.01.016

Legendre, M., Bartoli, J., Shmakova, L., Jeudy, S., Labadie, K., Adrait, A., et al. (2014). Thirty-thousand-year-old distant relative of giant icosahedral DNA viruses with a pandoravirus morphology. *Proc. Natl. Acad. Sci. U S A.* 111:4274. doi: 10.1073/pnas.1320670111

Legendre, M., Lartigue, A., Bertaux, L., Jeudy, S., Bartoli, J., Lescot, M., et al. (2015). In-depth study of Mollivirus sibericum, a new 30,000-y-old giant virus infecting *Acanthamoeba*. *Proc. Natl. Acad. Sci. U S A.* 112, E5327–E5335. doi: 10.1073/pnas.1510795112

Li, J. L., Cornman, R. S., Evans, J. D., Pettis, J. S., Zhao, Y., Murphy, C., et al. (2014). Systemic spread and propagation of a plant-pathogenic virus in European honeybees. *Apis mellifera*. *mBio* 5, e00898-13. doi: 10.1128/mBio.00898-813

Li, W., and Godzik, A. (2006). Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* 22, 1658–1659. doi: 10.1093/bioinformatics/btl158

Lwoff, A. (1957). The concept of virus. *Microbiology* 17, 239–253. doi: 10.1099/00221287-17-2-239

Macera, L., Antonelli, G., Spezia, P. G., Focosi, D., and Maggi, F. (2019). Lack of *Marseillevirus* DNA in immunocompetent and immunocompromised Italian patients. *J. Med. Virol.* 92, 187–190. doi: 10.1002/jmv.25592

Moreira, D., and López-García, P. (2009). Ten reasons to exclude viruses from the tree of life. *Nat. Rev. Microbiol.* 7:306. doi: 10.1038/nrmicro2108

Mosteo, R., Goñi, P., Miguel, N., Abadías, J., Valero, P., and Ormad, M. P. (2016). Bioaccumulation of pathogenic bacteria and amoeba by zebra mussels and their presence in watercourses. *Environ. Sci. Pollut. Res.* 23, 1833–1840. doi: 10.1007/s11356-015-5418-5412

Mueller, L., Baud, D., Bertelli, C., and Greub, G. (2013). *Lausannevirus* seroprevalence among asymptomatic young adults. *Intervirology* 56, 430–433. doi: 10.1159/000354565

Muli, E., Patch, H., Frazier, M., Frazier, J., Torto, B., Baumgarten, T., et al. (2014). Evaluation of the distribution and impacts of parasites, pathogens, and pesticides on honey bee (Apis mellifera) populations in East Africa. *PLoS One* 9:e94459. doi: 10.1371/journal.pone.0094459

Oliveira, G., Silva, L., Leão, T., Mougari, S., da Fonseca, F. G., Kroon, E. G., et al. (2019). Tupanvirus-infected amoebas are induced to aggregate with uninfected cells promoting viral dissemination. *Sci. Rep.* 9:183. doi: 10.1038/s41598-018-36552-36554

Oliveira, G. P., Lima, M. T., Arantes, T. S., Assis, F. L., Rodrigues, R. A. L., da Fonseca, F. G., et al. (2017). The investigation of promoter sequences of *Marseilleviruses* highlights a remarkable abundance of the AAATATTT motif in intergenic regions. *J. Virol.* 91:e1088-17. doi: 10.1128/JVI.01088-1017

Otta, D. A., Rott, M. B., Carlesso, A. M., and da Silva, O. S. (2012). Prevalence of *Acanthamoeba spp*. (Sarcomastigophora: Acanthamoebidae) in

wild populations of *Aedes aegypti* (Diptera: Culicidae). *Parasitol. Res.* 111, 2017–2022. doi: 10.1007/s00436-012-3050-3053

Pagnier, I., Reteno, D. G. I., Saadi, H., Boughalmi, M., Gaia, M., Slimani, M., et al. (2013). A decade of improvements in mimiviridae and marseilleviridae isolation from amoeba. *Intervirology* 56, 354–363. doi: 10.1159/000354556

Philippe, N., Legendre, M., Doutre, G., Couté, Y., Poirot, O., Lescot, M., et al. (2013). Pandoraviruses: amoeba viruses with genomes up to 2.5 Mb reaching that of parasitic eukaryotes. *Science* 341:281. doi: 10.1126/science.1239181

Popgeorgiev, N., Boyer, M., Fancello, L., Monteil, S., Robert, C., Rivet, R., et al. (2013a). Marseillevirus-Like virus recovered from blood donated by asymptomatic humans. *J. Infect. Dis.* 208, 1042–1050. doi: 10.1093/infdis/jit292

Popgeorgiev, N., Colson, P., Thuret, I., Chiarioni, J., Gallian, P., Raoult, D., et al. (2013b). *Marseillevirus* prevalence in multitransfused patients suggests blood transmission. *J. Clin. Virol.* 58, 722–725. doi: 10.1016/j.jcv.2013.10.001

Popgeorgiev, N., Michel, G., Lepidi, H., Raoult, D., and Desnues, C. (2013c). *Marseillevirus adenitis* in an 11-month-old child. *J. Clin. Microbiol.* 51, 4102–4105. doi: 10.1128/JCM.01918-1913

Popper, K. (2005). *The Logic of Scientific Discovery*. Milton Park: Routledge.

Rampelli, S., Soverini, M., Turroni, S., Quercia, S., Biagi, E., Brigidi, P., et al. (2016). ViromeScan: a new tool for metagenomic viral community profiling. *BMC Genomics* 17:165. doi: 10.1186/s12864-016-2446-3

Raoult, D. (2013). TRUC or the need for a new microbial classification. *Intervirology* 56, 349–353. doi: 10.1159/000354269

Raoult, D., Audic, S., Robert, C., Abergel, C., Renesto, P., Ogata, H., et al. (2004). The 1.2-megabase genome sequence of *Mimivirus*. *Science* 306, 1344–1350. doi: 10.1126/science.1101485

Raoult, D., and Boyer, M. (2010). Amoebae as genitors and reservoirs of giant viruses. *Intervirology* 53, 321–329. doi: 10.1159/000312917

Raoult, D., and Forterre, P. (2008). Redefining viruses: lessons from *Mimivirus*. *Nat. Rev. Microbiol.* 6, 315–319. doi: 10.1038/nrmicro1858

Raoult, D., La Scola, B., and Birtles, R. (2007). The discovery and characterization of *Mimivirus*, the largest known virus and putative pneumonia agent. *Clin. Infect. Dis.* 45, 95–102. doi: 10.1086/518608

Reteno, D. G., Benamar, S., Khalil, J. B., Andreani, J., Armstrong, N., Klose, T., et al. (2015). Faustovirus, an asfarvirus-Related new lineage of giant viruses infecting amoebae. *J. Virol.* 89:6585. doi: 10.1128/JVI.00115-115

Rodrigues, R. A. L., Louazani, A. C., Picorelli, A., Oliveira, G. P., Lobo, F. P., Colson, P., et al. (2020). Analysis of a *Marseillevirus* transcriptome reveals temporal gene expression profile and host transcriptional shift. *Front. Microbiol.* 11:651. doi: 10.3389/fmicb.2020.00651

Rolland, C., Andreani, J., Cherif Louazani, A., Aherfi, S., Francis, R., Rodrigues, R., et al. (2019). Discovery and further studies on giant viruses at the IHU mediterranee infection that modified the perception of the virosphere. *Viruses* 11:312. doi: 10.3390/v11040312

Saadi, H., Pagnier, I., Colson, P., Cherif, J. K., Beji, M., Boughalmi, M., et al. (2013). First isolation of *Mimivirus* in a patient with pneumonia. *Clin. Infect. Dis.* 57, 127–134. doi: 10.1093/cid/cit354

Sauvage, V., Livartowski, A., Boizeau, L., Servant-Delmas, A., Lionnet, F., Lefrère, J.-J., et al. (2014). No evidence of *Marseillevirus*-like virus presence in blood donors and recipients of multiple blood transfusions. *J. Infect. Dis.* 210, 2017–2018. doi: 10.1093/infdis/jiu443

Schulz, F., Alteio, L., Goudeau, D., Ryan, E. M., Yu, F. B., Malmstrom, R. R., et al. (2018). Hidden diversity of soil giant viruses. *Nat. Commun.* 9:4881. doi: 10.1038/s41467-018-07335-7332

Schulz, F., Yutin, N., Ivanova, N. N., Ortega, D. R., Lee, T. K., Vierheilig, J., et al. (2017). Giant viruses with an expanded complement of translation system components. *Science* 356:82. doi: 10.1126/science.aal4657

Schulz, F., Andreani, J., Francis, R., Boudjemaa, H., Khalil, J. Y. B., Lee, J., et al. (2020). Advantages and limits of metagenomic assembly and binning of a giant virus. *mSystems* 5:e00048-20. doi: 10.1128/mSystems.00048-20

Sharma, V., Colson, P., Pontarotti, P., and Raoult, D. (2016). Mimivirus inaugurated in the 21st century the beginning of a reclassification of viruses. *Curr. Opin. Microbiol.* 31, 16–24. doi: 10.1016/j.mib.2015.12.010

Silva, L. C. F., Almeida, G. M. F., Oliveira, D. B., Dornas, F. P., Campos, R. K., La Scola, B., et al. (2014). A resourceful giant: APMV is able to interfere with the human type I interferon system. *Microbes Infect.* 16, 187–195. doi: 10.1016/j.micinf.2013.11.011

Sobhy, H., La Scola, B., Pagnier, I., Raoult, D., and Colson, P. (2015). Identification of giant Mimivirus protein functions using RNA interference. *Front. Microbiol.* 6:345. doi: 10.3389/fmicb.2015.00345

Stasiak, K., Renault, S., Demattei, M. V., Bigot, Y., and Federici, B. A. (2003). Evidence for the evolution of ascoviruses from iridoviruses. *J. Gen. Virol.* 84, 2999–3009. doi: 10.1099/vir.0.19290-19290

Suzan-Monti, M., La Scola, B., and Raoult, D. (2006). Genomic and evolutionary aspects of *Mimivirus*. *Virus Res.* 117, 145–155. doi: 10.1016/j.virusres.2005.07.011

Takemura, M. (2016). Draft genome sequence of tokyovirus, a member of the family *Marseilleviridae* isolated from the Arakawa river of Tokyo. Japan. *Genome Announc.* 4:e00429-16. doi: 10.1128/genomeA.00429-416

Tan, Y. F., Lim, C. Y., Chong, C. W., Lim, P. K. C., Yap, I. K. S., Leong, P. P., et al. (2018). Isolation and quantification of *Mimivirus*-Like and *Marseillevirus*-Like viruses from soil samples in an aboriginal (*Orang asli*) village in Peninsular Malaysia. *Intervirology* 61, 92–95. doi: 10.1159/000491602

Thomas, V., Bertelli, C., Collyn, F., Casson, N., Telenti, A., Goesmann, A., et al. (2011). *Lausannevirus*, a giant amoebal virus encoding histone doublets. *Environ. Microbiol.* 13, 1454–1466. doi: 10.1111/j.1462-2920.2011.02446.x

Thomas, V., Loret, J.-F., Jousset, M., and Greub, G. (2008). Biodiversity of amoebae and amoebae-resisting bacteria in a drinking water treatment plant. *Environ. Microbiol.* 10, 2728–2745. doi: 10.1111/j.1462-2920.2008.01693.x

Van Etten, J. L. (2003). Unusual life style of giant chlorella viruses. *Annu. Rev. Genet.* 37, 153–195. doi: 10.1146/annurev.genet.37.110801.143915

Walker, P. J., Siddell, S. G., Lefkowitz, E. J., Mushegian, A. R., Dempsey, D. M., Dutilh, B. E., et al. (2019). Changes to virus taxonomy and the international code of virus classification and nomenclature ratified by the international committee on taxonomy of viruses (2019). *Arch. Virol.* 164, 2417–2429. doi: 10.1007/s00705-019-04306-w

Wardley, R. C., de, M., Andrade, C., Black, D. N., de Castro Portugal, F. L., Enjuanes, L., et al. (1983). African swine fever virus. *Brief review. Arch. Virol.* 76, 73–90. doi: 10.1007/BF01311692

Weisman, R. A., and Korn, E. D. (1967). Phagocytosis of latex beads by acanthamoeba. i. biochemical properties *. *Biochemistry* 6, 485–497. doi: 10.1021/bi00854a017

Yutin, N., and Koonin, E. V. (2012). Hidden evolutionary complexity of nucleo-cytoplasmic large DNA viruses of eukaryotes. *Virol. J.* 9:161. doi: 10.1186/1743-422X-9-161

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Microbiology

ORIGINAL RESEARCH
published: 07 June 2021
doi: 10.3389/fmicb.2021.683294

Discovery of Viral Myosin Genes With Complex Evolutionary History Within Plankton

*Soichiro Kijima[1], Tom O. Delmont[2], Urara Miyazaki[1,3], Morgan Gaia[2], Hisashi Endo[1] and Hiroyuki Ogata[1]**

[1] Chemical Life Science, Institute for Chemical Research, Kyoto University, Uji, Japan, [2] Metabolic Genomics, Genoscope, Institut de Biologie François Jacob, CEA, CNRS, Univ Evry, Université Paris Saclay, Évry-Courcouronnes, France, [3] Laboratory of Marine Environmental Microbiology, Division of Applied Biosciences, Graduate School of Agriculture, Kyoto University, Kyoto, Japan

Nucleocytoplasmic large DNA viruses (NCLDVs) infect diverse eukaryotes and form a group of viruses with capsids encapsulating large genomes. Recent studies are increasingly revealing a spectacular array of functions encoded in their genomes, including genes for energy metabolisms, nutrient uptake, as well as cytoskeleton. Here, we report the discovery of genes homologous to myosins, the major eukaryotic motor proteins previously unrecognized in the virosphere, in environmental genomes of NCLDVs from the surface of the oceans. Phylogenetic analyses indicate that most viral myosins (named "virmyosins") belong to the *Imitervirales* order, except for one belonging to the *Phycodnaviridae* family. On the one hand, the phylogenetic positions of virmyosin-encoding *Imitervirales* are scattered within the *Imitervirales*. On the other hand, *Imitervirales* virmyosin genes form a monophyletic group in the phylogeny of diverse myosin sequences. Furthermore, phylogenetic trends for the virmyosin genes and viruses containing them were incongruent. Based on these results, we argue that multiple transfers of myosin homologs have occurred not only from eukaryotes to viruses but also between viruses, supposedly during co-infections of the same host. Like other viruses that use host motor proteins for their intracellular transport or motility, these viruses may use the virally encoded myosins for the intracellular trafficking of giant viral particles.

Keywords: NCLDV, giant viruses, myosin, phylogeny, viral diversity, *Nucleocytoviricota*

OPEN ACCESS

Edited by:
Jonatas Abrahao, Federal University of Minas Gerais, Brazil

Reviewed by:
Rodrigo Araújo Lima Rodrigues, Federal University of Minas Gerais, Brazil
Philippe Colson, IHU Mediterranee Infection, France

****Correspondence:***
Hiroyuki Ogata
ogata@kuicr.kyoto-u.ac.jp

Specialty section:
This article was submitted to Virology, a section of the journal Frontiers in Microbiology

Received: *20 March 2021*
Accepted: *12 May 2021*
Published: *07 June 2021*

Citation:
Kijima S, Delmont TO, Miyazaki U, Gaia M, Endo H and Ogata H (2021) Discovery of Viral Myosin Genes With Complex Evolutionary History Within Plankton. Front. Microbiol. 12:683294. doi: 10.3389/fmicb.2021.683294

INTRODUCTION

Viruses were considered as tiny and simple biological objects until La Scola et al. (2003) discovered a giant virus from the water of a cooling tower. The virus named mimivirus is 750 nm in particle size and possesses a 1,182 kbp genome (Colson et al., 2017), a dimension that was large and complex enough to blow off the classical perception of viruses. After the discovery of mimivirus, related viruses were isolated including marseilleviruses, pandoraviruses, and pithoviruses, many of them with similar or even larger-sized particles or genomes (Abergel et al., 2015; Colson et al., 2017). These giant viruses infect diverse eukaryotes, possess a double-stranded DNA genome, belong to the phylum *Nucleocytoviricota* (Koonin et al., 2020), and are commonly referred to as

nucleocytoplasmic large DNA viruses (NCDLVs) (Iyer et al., 2006, p. 200). The monophyletic origin of NCLDVs has been suggested based on the presence of about 40 core genes of NCLDVs that can be traced back to their putative last common ancestor (Koonin and Yutin, 2019) as well as the congruent phylogenies of the most conserved eight proteins responsible for virion morphogenesis and informational processes (Guglielmini et al., 2019).

Because of their large virions, NCLDVs can encapsulate a large genome (several hundred kb up to 2.5 Mb) in their particles. Smaller viruses (such as small RNA viruses) encode only genes that are essential for their genome replication and capsid formation, while NCLDVs encode numerous genes that are not directly involved in their genome replication and virion morphogenesis (Moniruzzaman et al., 2020). These genes, often called auxiliary metabolic genes, are considered to function in reprogramming host metabolism and molecular machinery during viral infection toward enhancing viral replication and subsequent transmission to another host. For example, the recently characterized Prymnesium kappa virus RF01 encodes genes for all four succinate dehydrogenase subunits, as well as genes for modulating β-oxidation pathway (Blanc-Mathieu et al., 2021). These viral genes are suggested to boost energy production during viral replication, which can deteriorate host metabolism, or to enhance the supply of building blocks for viral replication. Another recent study reported the presence of actin genes (viractins) in NCLDV genomes (Cunha et al., 2020). Viractins are hypothesized to help viral infections by controlling the localization of the viral factory close to the host nucleus.

Hundreds of genomes have already been sequenced for cultured NCLDVs, yet these represent only the tip of iceberg of the diverse NCLDVs uncovered through environmental surveys (Schulz et al., 2020). To bypass cultivation, genome-resolved metagenomics has been applied to large metagenomic surveys, including on oceanic samples collected by *Tara* Oceans (Sunagawa et al., 2020), in order to characterize NCLDV metagenome-assembled genomes (MAGs) containing the gene pool of thousands of those viruses (Moniruzzaman et al., 2020; Schulz et al., 2020). NCLDV MAGs revealed a cosmopolitan nature of these viruses, extensive gene transfers with eukaryotes, and their complex metabolic capabilities.

In this study, we describe the identification of myosin genes in previously published NCLDV MAGs as well as newly identified ones derived from a manual binning and curation effort focused on large cellular size fractions of *Tara* Oceans enriched in NCLDVs when infecting planktonic eukaryotes (see section "Materials and Methods"). Myosin genes, which have not been previously described in viral genomes from cultures, form a superfamily of motor proteins involved in a wide range of motility processes in eukaryotic cells. Myosins have been grouped into various classes (Odronitz and Kollmar, 2007). Most myosins are classified into class 1–35 based on their phylogenetic relationships, while other myosins are phylogenetically orphan and classified into class A to U (Odronitz and Kollmar, 2007). The functions of orphan myosins are often unknown, while the functions of some members of the class 1–35 are characterized. For example, some myosins of class 2 contract muscle (Pertici et al., 2018), while some myosins of class 5 transport specific material along actin filaments (Hammer and Sellers, 2011). The various functions of myosins are supported by the head domains (Odronitz and Kollmar, 2007), which are universally conserved among myosins, interact with actins, and frequently serve for phylogenetic analyses (Odronitz and Kollmar, 2007). The head domains of myosins interact with actins when myosins bind ADP, whereas, when ADP is absent or ATP is in the ADP-binding site of myosins, the head domains no longer interact with actins. The tail part contains various domains, including coiled-coil domains (class 2 myosins), ankyrin repeats, and IQ motifs (Odronitz and Kollmar, 2007).

MATERIALS AND METHODS

NCLDV MAGs Derived From the Tara Ocean Project

Newly identified NCLDV MAGs were manually characterized and curated from the *Tara* Oceans metagenomes (size fractions > 0.8 μm), based on an initial binning strategy at large-scale focused on eukaryotes (Delmont et al., 2020), and following the same workflow as in previous studies (Cunha et al., 2020; Delmont et al., 2020; Kaneko et al., 2021). Briefly, metagenomes were organized into 11 sets based on their geography, and each set was co-assembled using MEGAHIT (Li et al., 2015) v.1.1.1. For each set, scaffolds longer than 2.5 kbp were processed within the bioinformatics platform anvi'o v.6.1 (Eren et al., 2015) to generate genome-resolved metagenomes (Delmont et al., 2018). CONCOCT (Alneberg et al., 2014) was used to identify large clusters of contigs within the set. We then used HMMER (Eddy, 2011) v3.1b2 to search for eight NCLDV gene markers (Guglielmini et al., 2019), and identified NCLDV MAGs by manually binning CONCOCT clusters using the anvi'o interactive interface. Finally, NCLDV MAGs were manually curated using the same interface, to minimize contamination as described previously (Delmont and Eren, 2016).

Sequence Datasets

To prepare a sequence set for the DNA polymerase elongation subunit family B (PolB), we first extracted PolB protein sequences from NCVOG (Yutin et al., 2009). Next, we collected PolB sequences by performing BLASTP from Virus-Host DB (Mihara et al., 2016) against the PolB sequences from NCVOG. We retained hits with an E-value < 1e-10. To identify PolBs in the NCLDV MAGs, we performed BLASTP from sequences derived from NCDLV MAGs generated by Moniruzzaman et al. (2020); Schulz et al. (2020), and ourselves (*Tara* Oceans MAGs) against the PolB sequences from NCVOG and Virus-Host DB. We retained hits with an E-value < 1e-10 and with their length in a range from 800 amino acids (aa) up to 1,800 aa to exclude anomalously short or long sequences that are difficult to align. We pooled these PolB sequences, and then removed redundancy using cd-hit (4.8.1) (Li and Godzik, 2006) and manually curated the dataset to reduce its size.

To identify eight core genes of NCLDVs [DNA polymerase elongation subunit family B, D5-like helicase-primase, Poxvirus

Late Transcription Factor VLTF3 like, Transcription factor S-II (TFIIS), packaging ATPase, NCLDV major capsid protein, DNA-directed RNA polymerase subunit alpha, and DNA-directed RNA polymerase subunit beta] (Koonin and Yutin, 2019) in NCLDV MAGs, we performed BLASTP (E-value < 1e-5, no length cutoff) from protein sequences derived from the MAGs against eight core genes recorded in NCVOG (Yutin et al., 2009).

For myosin homologs, we used full-length myosin sequences from a previous study (Odronitz and Kollmar, 2007) as primal references for myosins. This dataset contains various classes of myosins from diverse organisms. Next, we extracted myosin homologs from MMETSP (Keeling et al., 2014) and RefSeq (O'Leary et al., 2016) by performing BLASTP (blast 2.11.0) against the primal references to generate secondary references for full-length myosin sequences. We considered hits with an E-value < 1e-10 in this search. We identified myosin head domains in the primal and secondary myosin reference sequences by performing hmmscan (HMMER 3.3.1) (Eddy, 2011). We used Pfam (El-Gebali et al., 2019) as the HMM model for the hmmscan search and considered hits that were annotated as "Myosin_head" and showed an E-value < 1e-10. From each taxonomic group (the rank just below "Eukaryota" in the NCBI taxonomy, mostly corresponding to a class or higher level) and for each represented myosin class, we retained one representative sequence from primary and secondary reference sequences. Thus, the selected non-redundant sequence set represents myosins from wide classes and wide taxonomic ranges. To identify myosin sequences in the NCLDV MAGs, we performed BLASTP from sequences derived from MAGs against the primal myosin references. We considered hits with an E-value < 1e-10. We identified myosin head domains by performing hmmscan (Eddy, 2011). We considered hits that were annotated as "Myosin_head" with an E-value < 1e-10 and length longer than 550 aa.

Multiple Sequence Alignment

Multiple sequence alignments were generated using MAFFT (v7.471) (Katoh et al., 2019) with the L-INS-i algorithm, which is suitable for sequences that have only one alignable domain. After multiple alignment, we removed gappy sites by using trimAl (v1.4.rev15) (Capella-Gutiérrez et al., 2009) with "-gappyout" option.

For myosin sequences from viruses and eukaryotes, we generated two trimmed alignments by removing gapped columns using "-strict" option (named "strict dataset"), in addition to the "-gappyout" option (named "gappy dataset"). Finally, we removed sequences which have more than 30% gaps along the entire length of the alignment from each dataset to exclude divergent sequences.

Phylogenetic Analysis

The phylogenetic analyses were conducted with the ML framework using IQ-TREE (1.6.12) (Minh et al., 2020). For each alignment, we used the best substitution model selected by ModelFinder (Kalyaanamoorthy et al., 2017) of IQ-TREE. The selected models are described in the figure legends. The branch support values were computed based on the non-parametric bootstrap method with 100 bootstrap replicates, as well as transfer bootstrap expectation (TBE) with BOOSTER-WEB (Lemoine et al., 2018). TBE replaces the branch presence frequency in non-parametric bootstrap proportion (i.e., the expectation of a 0/1 function) by the expectation of a nearly continuous function. TBE supports acknowledge the presence of unstable taxa (single taxa that tends to move in and out of clades) and provide an alternative view of the robustness of a tree.

In the reconstruction of the phylogenetic tree of PolBs, *Poxviridae* sequences have been shown to be difficult to confidently position and tend to reduce the global resolution of phylogenetic trees, and were therefore subsequently removed as suggested by Guglielmini et al. (2019) to enhance the resolution of the phylogenetic trees.

For myosin sequences from viruses and eukaryotes, we also built phylogenetic trees using RAxML (8.2.12) (Stamatakis, 2014) on both "gappy" and "strict" datasets. We tried four substitution models (i.e., PROTCATAUTO, PROTCATIAUTO, PROTGAMMAAUTO, and PROTGAMMAIAUTO) and selected the tree with the highest likelihood score. Therefore, we obtained four trees for these myosin sequences (i.e., gappy/IQ-TREE, strict/IQ-TREE, gappy/RAxML, and strict/RAxML).

Visualization

The phylogenetic trees were visualized with iTOL v4 (Letunic and Bork, 2019).

Motif Search

We performed InterProScan (5.44-79.0) (Mitchell et al., 2019) to find motifs against each protein sequence derived from the NCLDV MAGs.

RESULTS AND DISCUSSION

Myosin Genes in NCLDV Genomes

We identified myosin-related genes in a total of 24 NCLDV MAGs (out of 2,275 considered in our survey) by performing BLASTP against reference myosin sequences compiled from a previous study (Odronitz and Kollmar, 2007), through our own independent effort from a parallel environmental genomic survey (Ha et al., 2021). These NCLDV MAGs were derived from the large cellular size fractions of *Tara* Oceans (n = 10) and studies by Moniruzzaman et al. (2020) (n = 5) and Schulz et al. (2020) (n = 9). All NCLDV MAGs but one originate from marine plankton samples (**Supplementary Table 1**). The size and number of genes in these 24 NCLDV MAGs ranged from 239 to 996 kb (633 kb on average) and 306 to 1,071 (643 on average), respectively (**Supplementary Tables 1**, **2**). The average length of the viral myosin sequences was 964 aa (702–1,076 aa). The viral myosin sequences contain the head domain (651 aa on average, from 583 to 754 aa) and the tail domain (265 aa on average, from 5 to 381 aa). The tail domain did not show the coiled-coil domain, which is seen in the class 2 myosin (functioning in muscle). Instead, the tail domains of the virmyosins in some cases (15 out of 24) contained the IQ calmodulin-binding motif, which functions to bind calmodulin

and is often seen in various classes of myosin (**Supplementary Table 3**) (Odronitz and Kollmar, 2007). An alignment of myosin head domain sequences from various organisms with these viral sequences establishes their strong homology (**Supplementary Figure 1**). Eight of 24 viral myosin genes were detected in the previously published marine metatranscriptomic data (Carradec et al., 2018) at >95% nt identify, suggesting their transcriptional activity in the ocean. The taxonomies of the closest homologs of other genes encoded together with the myosin homologs in the same contigs (**Figure 1**) revealed that a large proportion of genes (up to 44% and 15% on average) in the contigs best match to NCLDV genes. Some of these contigs encode NCLDV marker genes (**Figure 1**). Furthermore, each contig contains a large proportion of putative genes without any significant hit in public databases. Most of the viral myosin genes show a relatively low G + C content (24.2% on average), although this was slightly lower than the remaining genic regions (28.1% on average) (**Supplementary Table 1**). These features are specific for NCLDVs, thus excluding the possibility of contaminated myosin homologs in the NCLDV MAGs for most cases. We thus designate the viral myosin homologs as virmyosins. However, this does not mean that all the analyzed 2,275 MAGs are free of contamination from other organisms or viruses. We also recognize that the evidence for the NCLDV origin is weak for short contigs even among those shown in **Figure 1**.

To investigate the evolutionary relationships between the virmyosin-encoding NCLDVs, we performed phylogenetic analyses based on DNA polymerases (PolBs), a commonly used phylogenetic maker for NCLDVs. Eighteen of the 24 virmyosin-harboring MAGs were found to also encode PolB and were thus subjected to this analysis. The generated tree indicates that all but one of these virmyosin-encoding MAGs belong to the *Imitervirales* order (including mimiviruses and alga-infecting mimivirus relatives), the grouping of which was supported with a bootstrap value of 100% (**Figure 2** and **Supplementary Figure 2**). However, the lineages of virmyosin-encoding MAGs were scattered within four distantly related clades in the *Imitervirales* branches (clades c1, c2, c3, and c4). Two of these clades (c3 and c4) were supported by a bootstrap value $\geq$ 91%, and all were supported by a TBE value $\geq$ 95%. The members of c1 and c3 are closely related to *Cafeteria roenbergensis* virus and *Aureococcus anophagefferens* virus, respectively. There were no closely related isolated viruses for c2 and c4. A previous study described the identification of viractin genes in two distant clades of the genomes of the *Imitervirales* order (Cunha et al., 2020). In the PolB tree we generated, the viractin-encoding MAGs were again placed within the *Imitervirales* branches but did not show close relationships with the virmyosin-encoding MAGs. One of the virmyosin-encoding MAGs was classified in the *Phycodnaviridae* family, being closely located to *Emiliania huxleyi* virus in the tree. This virmyosin was encoded in a contig harboring seven genes (MAG: TARA_ARC_NCLDV_Bin_278_6; contig: c_000000000006), of which two best hit to NCLDVs (**Figure 1**). Given the small size of the contig, the possibility of contamination cannot be totally excluded for this contig.

Six of the 24 virmyosin-harboring MAGs did not encode PolB. When searched against reference isolated viral genomes, most of the NCLDV core genes from these six MAGs showed the largest sequence similarities to the viruses of *Imitervirales*, suggesting that they are the members of this order (**Supplementary Table 4**).

Since myosins are universal in the eukaryotic domain, we hypothesized that these NCLDVs acquired myosin homologs by horizontal gene transfers (HGTs) from various eukaryotes. To determine the source eukaryotic lineages for the putative HGTs, we performed phylogenetic analyses on the virmyosin and reference myosin sequences. We generated four phylogenetic trees with different methods (**Figure 3** and

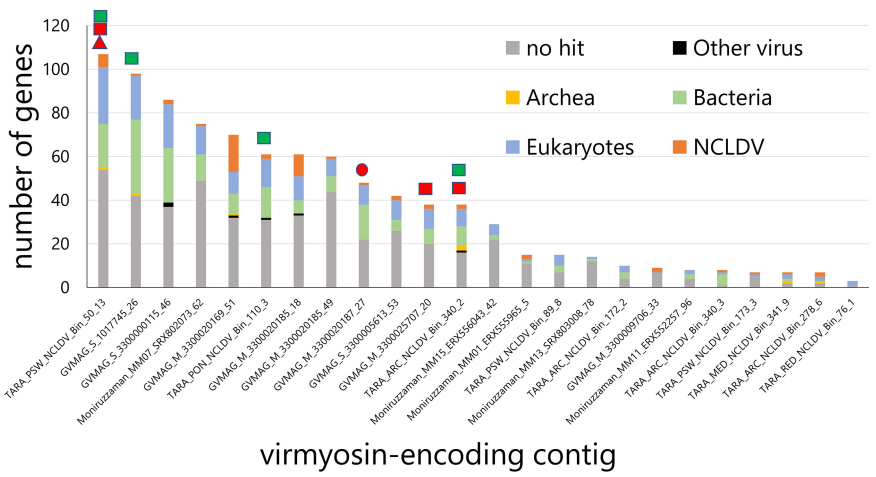


FIGURE 1 | Taxonomic annotation of genes encoded in the virmyosin-harboring NCLDV contigs. Best hit-based taxonomic annotation was performed for each gene using BLAST against the RefSeq database. Identifiers on the *x*-axis represent the names of MAGs that contain virmyosin-encoding contigs. Marks above the bars indicate the presence of NCLDV core genes in the corresponding contigs, including Poxvirus Late Transcription Factor VLTF3 like (red triangle), major capsid protein (red circle), RNA polymerase subunit alpha (red rectangle), and RNA polymerase subunit beta (green rectangle).

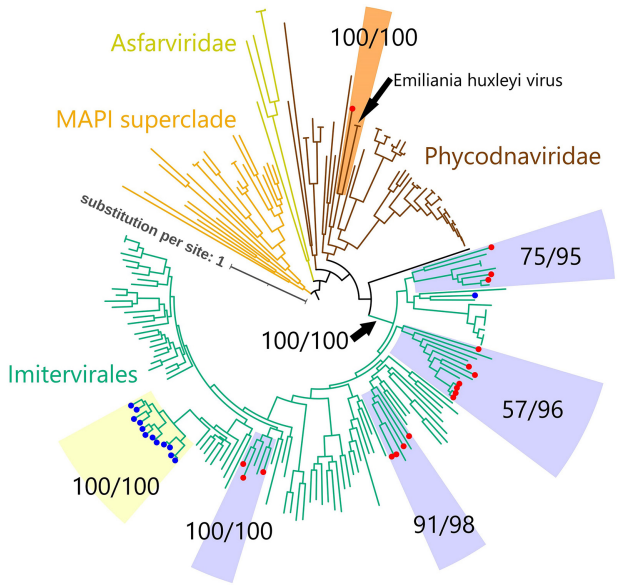


FIGURE 2 | Phylogenetic tree of NCLDVs based on DNA polymerase B. The tree was built from an alignment (993 sites) of 205 PolB sequences of NCLDVs (18 MAGs encoding myosin gene, 13 MAGs encoding actin gene, and 174 other broad and unbiased taxa from the reference database). Branches labeled with red and blue circles represent NCLDV MAGs with virmyosin and viractin genes, respectively. Branches are color-coded based on family- or order-level taxonomy. The MAPI superclade (shown in orange) comprises *Marseilleviridae*, *Ascoviridae*, Pitho-like viruses, and *Iridoviridae*. Paired-numerical values represent the non-parametric bootstrap (left) and TBE (right) values for the branch support. *Phycodnaviridae* MAG with virmyosin is phylogenetically close to *Emiliania huxleyi* virus with maximum bootstrap and TBE supports. We used MAPI superclade as an outgroup to root the tree. The LG + F + R10 substitution model was selected by IQ-TREE for the best model for tree reconstruction.

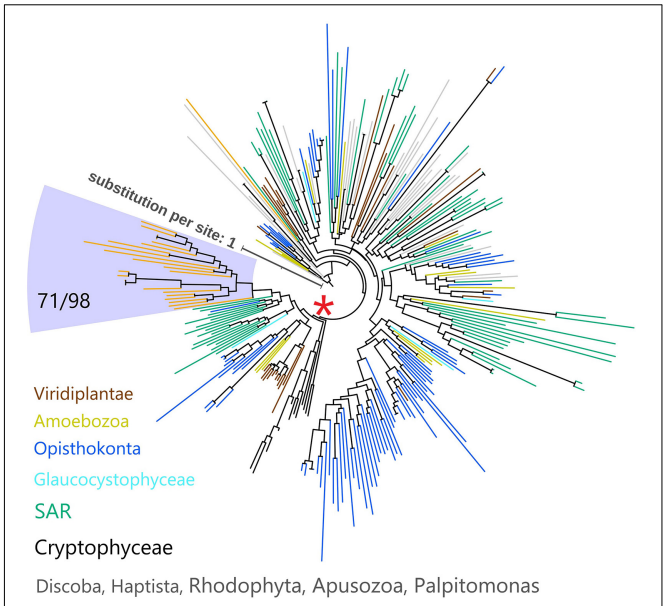


FIGURE 3 | Phylogeny of myosin from NCLDVs and eukaryotic lineages. The tree was built from the multiple sequence alignment (658 sites) of 286 myosin head domain sequences. Branches are color-coded based on taxonomic groups (orange branches correspond to virmyosins). The *Imitervirales* virmyosin clade (highlighted with a pale blue-colored background) was supported with a bootstrap value of 71% of and TBE of 98%. We consider myosin-II as an outgroup to root this tree. The LG + R10 substitution model was selected by IQ-TREE for the best model for tree reconstruction.

Supplementary Figures 3A–E; see section "Materials and Methods"). The phylogenetic placement of the virmyosin from the MAG belonging to the *Phycodnaviridae* family was unstable. However, all the trees showed a monophyletic group of *Imitervirales* virmyosins. The grouping is supported by a bootstrap value of 71% and TBE of 98% (**Figure 3** and **Supplementary Figure 3E**).

To examine the effect of long-branch attraction on the formation of the *Imitervirales* virmyosin clade, we computed the pair-wise sequence identities within virmyosins and between virmyosins and eukaryotic myosins (**Figure 4**). The sequence similarities for many of the virmyosin pairs were higher than those between virmyosins and eukaryotic myosins, thus diminishing the possibility of long branch attraction effect on the monophyletic grouping the *Imitervirales* virmyosins. The *Imitervirales* virmyosin clade is branching out from orphan myosins of the SAR (Stramenopiles, Alveolates, and Rhizaria) supergroup, but the grouping of these virmyosins and the orphan myosins was not supported. Finally, to further improve the tree reconstruction, we built a phylogenetic tree using the virmyosin sequences as well as eukaryotic myosin sequences that formed a clade together with the virmyosins (the clade marked with red "*" in **Figure 3**). The newly generated tree again displayed the monophyletic grouping of the *Imitervirales* virmyosins (bootstrap value, 98% and TBE, 100%) and placed it within a clade of myosin sequences from Apicomplexa (*Toxoplasma gondii*), Stramenopiles (the diatom *Thalassiosira pseudonana*, and the oomycete *Hyaloperonospora parasitica*), and a fungus (*Rhizopogon burlinghamii*) (**Figure 5** and **Supplementary Figures 3F,G**). We further regenerated a myosin tree by adding diverse eukaryotic myosin sequences (**Supplementary Figures 3H,I**). The tree showed a monophyletic group of the virmyosins within a clade enriched with sequences from SAR. However, the grouping of the virmyosins and eukaryotic sequences was not supported. This grouping and the fact that eukaryotes belonging to the SAR supergroup represent a major group of marine plankton suggest a possible SAR origin for virmyosins but leave the specific source lineage unresolved due to the lack of statistical support.

Possible HGT Scenarios for the Virmyosin

We next examined the congruence of the tree of virmyosins and the viral trees based on PolB and eight core genes of NCLDVs. In this analysis, we focused on 18 MAGs that encode both virmyosin and PolB. Congruence of the topologies between these trees (the virmyosin tree and viral trees) is expected if the original virmyosin gene was acquired by an ancestral virus prior to the divergence of viral families or orders. We

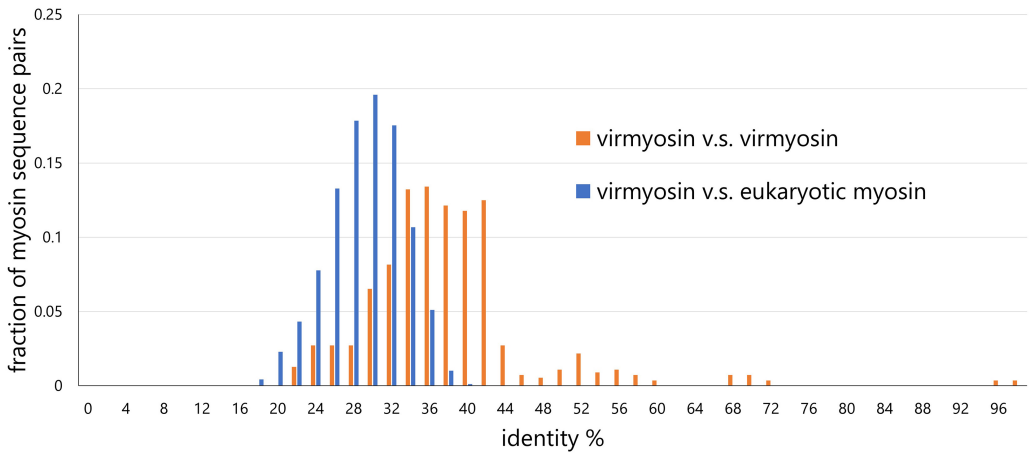


FIGURE 4 | Distributions of pair-wise myosin sequence similarities within the NCLDVs (virmyosin *vs* virmyosin) and between NCLDVs and eukaryotes (virmyosin *vs* eukaryotic myosin). This graph shows that the mean value of myosin similarity within NCLDVs is higher than that between NCLDVs and eukaryotes.

extracted subtrees for 18 virmyosins from **Figure 3** and 18 PolBs from **Figure 2** (**Figures 6A,B**). The tree based on the NCLDV core genes was newly reconstructed (**Figure 6C**), which was mostly congruent with the tree based on PolB (**Figure 6B**). To investigate the congruence of the trees, we focused on clades in the virmyosin tree that were supported with a bootstrap value greater than 99%. The analysis revealed the clades II and III form monophyletic groups also in the viral trees. However, the monophyletic clades I and IV in the virmyosin tree turned out to be polyphyletic in the viral trees with statistical supports (**Figure 6**). This result suggests that multiple and distantly related ancestral viruses of *Imitervirales* independently acquired myosin genes through HGTs.

The monophyletic grouping of the *Imitervirales* virmyosins and the scattered distribution of the virmyosin-encoding *Imitervirales* in the viral trees are intriguing. We consider four major possibilities for the origin of virmyosins. The first possibility is that all the *Imitervirales* myosin genes were vertically inherited from an ancestral virus of the *Imitervirales*, and subsequently lost independently in most of the descendant lineages. This scenario is not parsimonious as it needs to assume many independent gene losses, given 24 virmyosin-encoding NCLDVs and over 2,000 NCLDVs lacking virmyosin. The second possibility is that myosin genes were independently recruited multiple times from eukaryotes by ancestral viruses belonging to the *Imitervirales*. This scenario can account for the topological differences between the virmyosin and viral trees. However, this scenario cannot readily explain the monophyletic grouping of the virmyosins, as this scenario needs to additionally assume independent horizontal acquisitions of myosin genes by distantly related ancestral viruses from the same or closely related eukaryotes (e.g., ancestral SAR). Such acquisitions seem implausible because host changes are likely rampant events for *Imitervirales* given the wide host ranges of the known viruses in this group (Sun et al., 2020). The third possibility is that a myosin gene transfer occurred once in a viral lineage of *Imitervirales* from its host. Then, after the viral myosin gene acquired beneficial functions for the virus, this gene was transferred to other viruses, probably during co-infection in the same host (which may be different from the current hosts). Of note, a previous study reported a clear case of HGT between NCLDVs (Christo-Foroux et al., 2020). This scenario can explain both the topological difference between the virmyosin and viral trees and the monophyletic grouping of

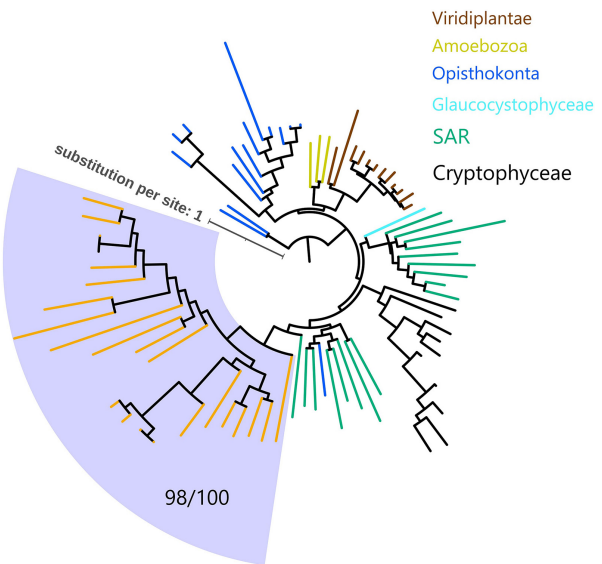


FIGURE 5 | Phylogeny of virmyosins and their close relatives in eukaryotes. The tree was built from an alignment (688 sites) of 81 myosin sequences of NCLDVs, Amoebozoa, SAR (Stramenopiles, Alveolata, and Rhizaria), Opisthokonta, Cryptophyceae, Viridiplantae, and Glaucocystophyceae. Background of the *Imitervirales* virmyosin clade is pale blue-colored. Branches are color-coded based on taxonomic groups (orange branches correspond to virmyosins). We consider myosin-II of *Nasonia vitripennis* and *Neurospora crassa* as outgroups to root this tree. The LG + F + R7 substitution model was selected by IQ-TREE for the best model for tree reconstruction.

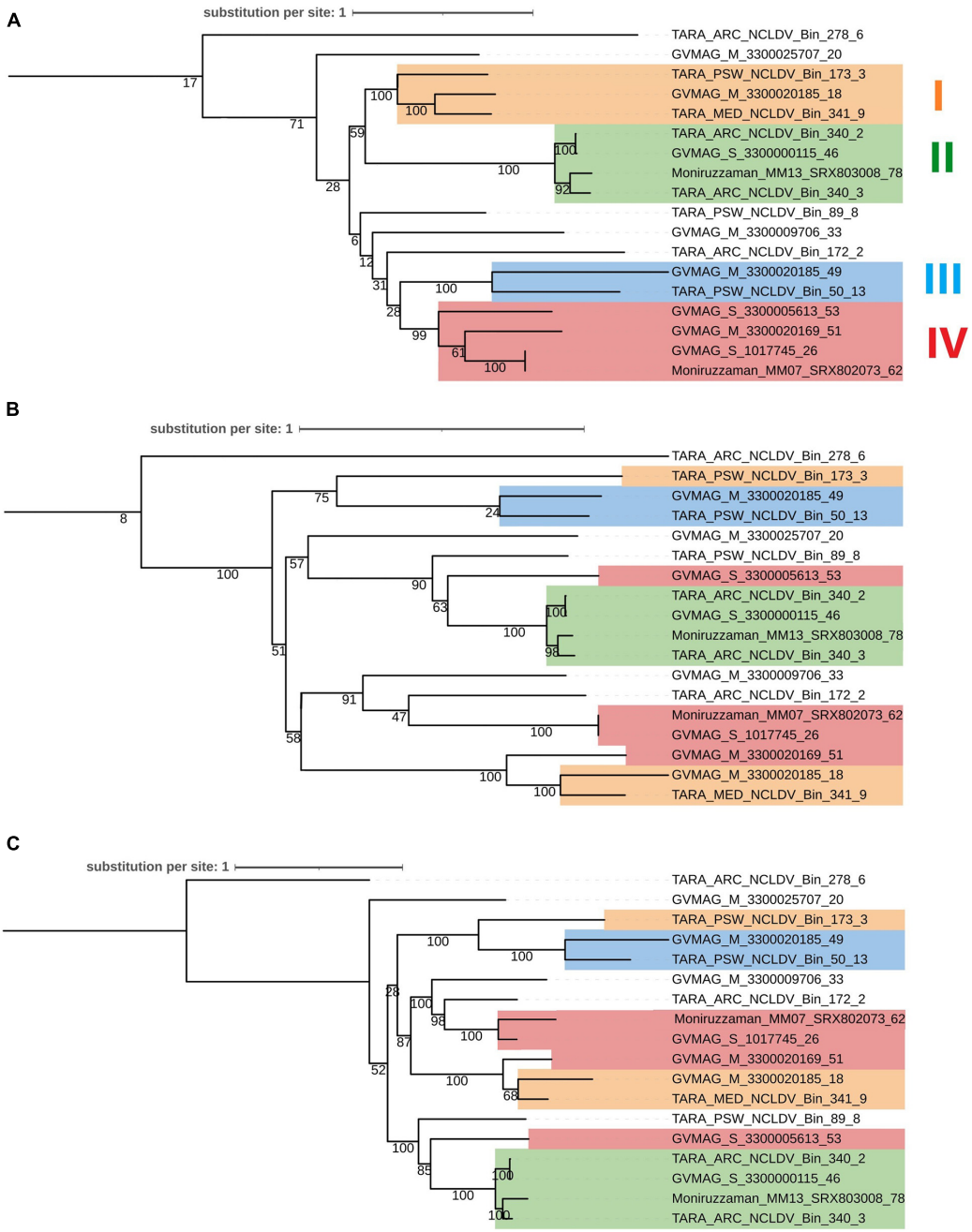


FIGURE 6 | Phylogenetic trees of **(A)** viral myosins, **(B)** PolBs, and **(C)** core gene sequences from MAGs encoding both genes myosin and PolB. Labels at the leaves are the identifiers of the NCLDV MAGs. Clades supported with a parametric bootstrap value > 99% in the myosin tree are marked in both trees by colored rectangles. This myosin and PolB trees were generated from **Figure 3** and **Figure 2** by extracting MAGs encoding both myosin and PolB, respectively. Clades I–IV correspond to highly supported groups in the myosin tree.

the virmyosins. Finally, as a remote possibility, myosin genes could have been transferred from an ancient giant virus to a eukaryotic ancestor [either before or after the Last Eukaryotic Common Ancestor (LECA)] as suggested for DNA-dependent RNA polymerases (Guglielmini et al., 2019). This scenario is not supported by the tree topology of myosins. However, given the unresolved deep branches in the tree especially for the virmyosin clade, this scenario cannot be excluded as a possible one. In this case, gene transfers between NCLDVs still need to be invoked to account for the monophyly of virmyosins. Further exploration of the actual diversity of *Imitervirales*, and more globally of NCLDVs, from various environments will certainly provide important insights regarding the robustness of these hypotheses.

Possible Roles of Virmyosin

Myosins are known to walk along actin filaments, which are located at the peripheric side of the cytoplasm (Geada et al., 2001; Handa et al., 2013). For example, the actin and microtubule cytoskeletons cooperate in organelle transport in a variety of situations (Goode et al., 2000; Rogers and Gelfand, 2000). Secretory vesicles also move by both microtubule- and actin-based molecular motors (Geada et al., 2001). Some classes of myosins function as a transporter to carry specific materials including viruses. For example, in the case of human immunodeficiency virus and influenza A virus, virus-containing endocytic viral cargoes propelled by class 2 myosin moving on filamentous actin spread to neighboring uninfected cells (Kadiu and Gendelman, 2011; Roberts et al., 2015). Motor proteins are used for the intracellular motility of NCLDVs. For example, motor proteins such as kinesin and dynein are known to transport African swine fever virus on microtubules (Jouvenet and Wileman, 2005). Intracellular enveloped viruses of vaccinia virus are transported toward the cell surface in a microtubule-dependent process (Hollinshead et al., 2001). Virmyosin identified in the environmental NCLDV genomes in this study may function in intracellular transport of virions like these viruses. Of note, there is no clear case for the use of myosin molecular motors by NCLDVs for their intracellular transport or motility. It has been speculated that the slow movements of intracellular enveloped virus of vaccinia virus that persist in the absence of microtubules may be related to the action of myosin motors (Geada et al., 2001). It is also known that the formation of actin tail, which propels virions of vaccinia virus, requires host Myosin-9A (Handa et al., 2013), but the role Myosin-9A as molecular motor in this process is unknown.

CONCLUSION

In this study, we provided strong evidence showing that marine members of NCLDVs encode myosin homologs (virmyosins). The function of virmyosin could not be inferred based on the similarity to functionally characterized myosin homologs. Together with the previously discovered actin homologs in NCLDVs, these results suggest that the genetic independence of NCLDVs from their hosts encompasses a wide-range of cellular processes, including intracellular trafficking as implied by our study and the translation process as considered from previous discoveries of translation genes in this group of viruses. Our phylogenetic analyses suggest a complex evolutionary origin of the virmyosin genes, which may involve not only HGTs from eukaryotes to NCLDVs but also intra-virus HGTs within NCLDVs. The functions encoded in the huge genetic pool of NCLDVs are revealing the amazingly diverse strategies to control host cellular processes in this diverse group of viruses. Scrutinizing available and newly generated environmental genomes will contribute to better characterizing the infection strategies of this fascinating group of viruses.

DATA AVAILABILITY STATEMENT

Tara Oceans raw read data analyzed in Delmont et al. (2020) and the present study are available at the EBI under project PRJEB402 (https://www.ebi.ac.uk/ena/browser/view/PRJEB402). All the NCLDV MAG data presented in this study, as well as associated data (e.g., sequence datasets, multiple sequence alignments, and phylogenetic trees) are available from ftp://ftp.genome.jp/pub/db/community/tara/Virmyosin/ and https://www.genome.jp/ftp/db/community/tara/Virmyosin/.

AUTHOR CONTRIBUTIONS

SK performed most of the bioinformatics analyses and wrote the initial version of the manuscript. TD performed metagenomic analyses. UM contributed to the discovery of viral myosin genes. MG contributed to the phylogenetic and evolutionary analyses. HO designed the project. HE and HO supervised the work. All authors contributed to the interpretation of the data and finalization of the manuscript, and approved the final manuscript.

FUNDING

This study was supported by JSPS KAKANHI (18H02279), Research Unit for Development of Global Sustainability, Kyoto University Research Coordination Alliance, and the Collaborative Research Program of the Institute for Chemical Research, Kyoto University (No. 2020-28).

ACKNOWLEDGMENTS

We thank the *Tara* Oceans consortium and the people and sponsors who supported *Tara* Oceans. *Tara* Oceans (that includes both the *Tara* Oceans and *Tara* Oceans Polar Circle expeditions) would not exist without the leadership of the Tara Expeditions Foundation and the continuous support of 23 institutes (https://oceans.taraexpeditions.org). This article is contribution number 117 of *Tara* Oceans. Computation time was provided by the SuperComputer System, Institute for Chemical Research, Kyoto University.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2021.683294/full#supplementary-material

REFERENCES

Abergel, C., Legendre, M., and Claverie, J.-M. (2015). The rapidly expanding universe of giant viruses: mimivirus, pandoravirus, pithovirus and mollivirus. *FEMS Microbiol. Rev.* 39, 779–796. doi: 10.1093/femsre/fuv037

Alneberg, J., Bjarnason, B. S., de Bruijn, I., Schirmer, M., Quick, J., Ijaz, U. Z., et al. (2014). Binning metagenomic contigs by coverage and composition. *Nat. Methods* 11, 1144–1146. doi: 10.1038/nmeth.3103

Blanc-Mathieu, R., Dahle, H., Hofgaard, A., Brandt, D., Ban, H., Kalinowski, J., et al. (2021). A persistent giant algal virus, with a unique morphology, encodes an unprecedented number of genes involved in energy metabolism. *J. Virol.* 95:e02446–20. doi: 10.1128/JVI.02446-20

Capella-Gutiérrez, S., Silla-Martínez, J. M., and Gabaldón, T. (2009). trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* 25, 1972–1973. doi: 10.1093/bioinformatics/btp348

Carradec, Q., Pelletier, E., Da Silva, C., Alberti, A., Seeleuthner, Y., Blanc-Mathieu, R., et al. (2018). A global ocean atlas of eukaryotic genes. *Nat. Commun.* 9:373. doi: 10.1038/s41467-017-02342-1

Christo-Foroux, E., Alempic, J.-M., Lartigue, A., Santini, S., Labadie, K., Legendre, M., et al. (2020). Characterization of mollivirus kamchatka, the first modern representative of the proposed molliviridae family of giant viruses. *J. Virol.* 94:e01997–19. doi: 10.1128/JVI.01997-19

Colson, P., La Scola, B., and Raoult, D. (2017). giant viruses of amoebae: a journey through innovative research and paradigm changes. *Annu. Rev. Virol.* 4, 61–85. doi: 10.1146/annurev-virology-101416-041816

Cunha, V. D., Gaia, M., Ogata, H., Jaillon, O., Delmont, T. O., and Forterre, P. (2020). Giant viruses encode novel types of actins possibly related to the origin of eukaryotic actin: the viractins. *bioRxiv* [Preprint] doi: 10.1101/2020.06.16.150565

Delmont, T. O., and Eren, A. M. (2016). Identifying contamination with advanced visualization and analysis practices: metagenomic approaches for eukaryotic genome assemblies. *Peer J.* 4:e1839. doi: 10.7717/peerj.1839

Delmont, T. O., Gaia, M., Hinsinger, D. D., Fremont, P., Guerra, A. F., Eren, A. M., et al. (2020). Functional repertoire convergence of distantly related eukaryotic plankton lineages revealed by genome-resolved metagenomics. *bioRxiv* [Preprint] doi: 10.1101/2020.10.15.341214

Delmont, T. O., Quince, C., Shaiber, A., Esen, ÖC., Lee, S. T., Rappé, M. S., et al. (2018). Nitrogen-fixing populations of Planctomycetes and Proteobacteria are abundant in surface ocean metagenomes. *Nat. Microbiol.* 3, 804–813. doi: 10.1038/s41564-018-0176-9

Eddy, S. R. (2011). Accelerated profile HMM searches. *PLoS Comput. Biol.* 7:e1002195. doi: 10.1371/journal.pcbi.1002195

El-Gebali, S., Mistry, J., Bateman, A., Eddy, S. R., Luciani, A., Potter, S. C., et al. (2019). The Pfam protein families database in 2019. *Nucleic Acids Res.* 47, D427–D432. doi: 10.1093/nar/gky995

Eren, A. M., Esen, ÖC., Quince, C., Vineis, J. H., Morrison, H. G., Sogin, M. L., et al. (2015). Anvi'o: an advanced analysis and visualization platform for 'omics data. *Peer J.* 3:e1319. doi: 10.7717/peerj.1319

Geada, M. M., Galindo, I., Lorenzo, M. M., Perdiguero, B., and Blasco, R. (2001). Movements of vaccinia virus intracellular enveloped virions with GFP tagged to the F13L envelope protein. *J. Gen. Virol.* 82, 2747–2760. doi: 10.1099/0022-1317-82-11-2747

Goode, B. L., Drubin, D. G., and Barnes, G. (2000). Functional cooperation between the microtubule and actin cytoskeletons. *Curr. Opin. Cell Biol.* 12, 63–71. doi: 10.1016/s0955-0674(99)00058-7

Guglielmini, J., Woo, A. C., Krupovic, M., Forterre, P., and Gaia, M. (2019). Diversification of giant and large eukaryotic dsDNA viruses predated the origin of modern eukaryotes. *Proc. Natl. Acad. Sci. U.S.A.* 116, 19585–19592. doi: 10.1073/pnas.1912006116

Ha, A. D., Moniruzzaman, M., and Aylward, F. O. (2021). High transcriptional activity and diverse functional repertoires of hundreds of giant viruses in a coastal marine system. *bioRxiv* [Preprint] doi: 10.1101/2021.03.08.434518

Hammer, J. A., and Sellers, J. R. (2011). Walking to work: roles for class V myosins as cargo transporters. *Nat. Rev. Mol. Cell Biol.* 13, 13–26. doi: 10.1038/nrm3248

Handa, Y., Durkin, C. H., Dodding, M. P., and Way, M. (2013). Vaccinia virus F11 promotes viral spread by acting as a PDZ-containing scaffolding protein to bind myosin-9A and inhibit RhoA signaling. *Cell Host Microbe* 14, 51–62. doi: 10.1016/j.chom.2013.06.006

Hollinshead, M., Rodger, G., Van Eijl, H., Law, M., Hollinshead, R., Vaux, D. J., et al. (2001). Vaccinia virus utilizes microtubules for movement to the cell surface. *J. Cell Biol.* 154, 389–402. doi: 10.1083/jcb.200104124

Iyer, L. M., Balaji, S., Koonin, E. V., and Aravind, L. (2006). Evolutionary genomics of nucleo-cytoplasmic large DNA viruses. *Virus Res.* 117, 156–184. doi: 10.1016/j.virusres.2006.01.009

Jouvenet, N., and Wileman, T. (2005). African swine fever virus infection disrupts centrosome assembly and function. *J. Gen. Virol.* 86, 589–594. doi: 10.1099/vir.0.80623-0

Kadiu, I., and Gendelman, H. E. (2011). Human immunodeficiency virus type 1 endocytic trafficking through macrophage bridging conduits facilitates spread of infection. *J. Neuroimmune Pharmacol.* 6, 658–675. doi: 10.1007/s11481-011-9298-z

Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K. F., von Haeseler, A., and Jermiin, L. S. (2017). ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat. Methods* 14, 587–589. doi: 10.1038/nmeth.4285

Kaneko, H., Blanc-Mathieu, R., Endo, H., Chaffron, S., Delmont, T. O., Gaia, M., et al. (2021). Eukaryotic virus composition can predict the efficiency of carbon export in the global ocean. *iScience* 24:102002. doi: 10.1016/j.isci.2020.102002

Katoh, K., Rozewicki, J., and Yamada, K. D. (2019). MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. *Brief. Bioinform.* 20, 1160–1166. doi: 10.1093/bib/bbx108

Keeling, P. J., Burki, F., Wilcox, H. M., Allam, B., Allen, E. E., Amaral-Zettler, L. A., et al. (2014). The Marine Microbial Eukaryote Transcriptome Sequencing Project (MMETSP): illuminating the functional diversity of eukaryotic life in the oceans through transcriptome sequencing. *PLoS Biol.* 12:e1001889. doi: 10.1371/journal.pbio.1001889

Koonin, E. V., and Yutin, N. (2019). Evolution of the large nucleocytoplasmic DNA viruses of eukaryotes and convergent origins of viral gigantism. *Adv. Virus Res.* 103, 167–202. doi: 10.1016/bs.aivir.2018.09.002

Koonin, E. V., Dolja, V. V., Krupovic, M., Varsani, A., Wolf, Y. I., Yutin, N., et al. (2020). Global organization and proposed megataxonomy of the virus world. *Microbiol. Mol. Biol. Rev.* 84, e00061–19. doi: 10.1128/MMBR.00061-19

La Scola, B., Audic, S., Robert, C., Jungang, L., de Lamballerie, X., Drancourt, M., et al. (2003). A giant virus in amoebae. *Science* 299:2033. doi: 10.1126/science.1081867

Lemoine, F., Domelevo Entfellner, J.-B., Wilkinson, E., Correia, D., Dávila Felipe, M., De Oliveira, T., et al. (2018). Renewing felsenstein's phylogenetic bootstrap in the era of big data. *Nature* 556, 452–456. doi: 10.1038/s41586-018-0043-0

Letunic, I., and Bork, P. (2019). Interactive Tree Of Life (iTOL) v4: recent updates and new developments. *Nucleic Acids Res.* 47, W256–W259. doi: 10.1093/nar/gkz239

Li, D., Liu, C.-M., Luo, R., Sadakane, K., and Lam, T.-W. (2015). MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics* 31, 1674–1676. doi: 10.1093/bioinformatics/btv033

Li, W., and Godzik, A. (2006). Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* 22, 1658–1659. doi: 10.1093/bioinformatics/btl158

Mihara, T., Nishimura, Y., Shimizu, Y., Nishiyama, H., Yoshikawa, G., Uehara, H., et al. (2016). Linking virus genomes with host taxonomy. *Viruses* 8:66. doi: 10.3390/v8030066

Minh, B. Q., Schmidt, H. A., Chernomor, O., Schrempf, D., Woodhams, M. D., von Haeseler, A., et al. (2020). IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Mol. Biol. Evol.* 37, 1530–1534. doi: 10.1093/molbev/msaa015

Mitchell, A. L., Attwood, T. K., Babbitt, P. C., Blum, M., Bork, P., Bridge, A., et al. (2019). InterPro in 2019: improving coverage, classification and access to protein sequence annotations. *Nucleic Acids Res.* 47, D351–D360. doi: 10.1093/nar/gky1100

Moniruzzaman, M., Martinez-Gutierrez, C. A., Weinheimer, A. R., and Aylward, F. O. (2020). Dynamic genome evolution and complex virocell metabolism of globally-distributed giant viruses. *Nat. Commun.* 11:1710. doi: 10.1038/s41467-020-15507-2

O'Leary, N. A., Wright, M. W., Brister, J. R., Ciufo, S., Haddad, D., McVeigh, R., et al. (2016). Reference sequence (RefSeq) database at NCBI: current status,

taxonomic expansion, and functional annotation. *Nucleic Acids Res.* 44, D733–D745. doi: 10.1093/nar/gkv1189

Odronitz, F., and Kollmar, M. (2007). Drawing the tree of eukaryotic life based on the analysis of 2,269 manually annotated myosins from 328 species. *Genome Biol.* 8:R196. doi: 10.1186/gb-2007-8-9-r196

Pertici, I., Bongini, L., Melli, L., Bianchi, G., Salvi, L., Falorsi, G., et al. (2018). A myosin II nanomachine mimicking the striated muscle. *Nat. Commun.* 9:3532. doi: 10.1038/s41467-018-06073-9

Roberts, K. L., Manicassamy, B., and Lamb, R. A. (2015). Influenza A virus uses intercellular connections to spread to neighboring cells. *J. Virol.* 89, 1537–1549. doi: 10.1128/JVI.03306-14

Rogers, S. L., and Gelfand, V. I. (2000). Membrane trafficking, organelle transport, and the cytoskeleton. *Curr. Opin. Cell Biol.* 12, 57–62. doi: 10.1016/s0955-0674(99)00057-5

Schulz, F., Roux, S., Paez-Espino, D., Jungbluth, S., Walsh, D. A., Denef, V. J., et al. (2020). Giant virus diversity and host interactions through global metagenomics. *Nature* 578, 432–436. doi: 10.1038/s41586-020-1957-x

Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30, 1312–1313. doi: 10.1093/bioinformatics/btu033

Sun, T.-W., Yang, C.-L., Kao, T.-T., Wang, T.-H., Lai, M.-W., and Ku, C. (2020). Host range and coding potential of eukaryotic giant viruses. *Viruses* 12:1337. doi: 10.3390/v12111337

Sunagawa, S., Acinas, S. G., Bork, P., Bowler, C., Tara Oceans Coordinators, Eveillard, D., et al. (2020). Tara oceans: towards global ocean ecosystems biology. *Nat. Rev. Microbiol.* 18, 428–445. doi: 10.1038/s41579-020-0364-5

Yutin, N., Wolf, Y. I., Raoult, D., and Koonin, E. V. (2009). Eukaryotic large nucleo-cytoplasmic DNA viruses: clusters of orthologous genes and reconstruction of viral genome evolution. *Virol. J.* 6:223. doi: 10.1186/1743-422X-6-223

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Microbiology

REVIEW
published: 22 June 2021
doi: 10.3389/fmicb.2021.701611

Main Targets of Interest for the Development of a Prophylactic or Therapeutic Epstein-Barr Virus Vaccine

*Vincent Jean-Pierre[1], Julien Lupo[1], Marlyse Buisson[1,2], Patrice Morand[1] and Raphaële Germi[1]**

[1] *Laboratoire de Virologie, Institut de Biologie et de Pathologie, CHU de Grenoble Alpes, Grenoble, France,* [2] *Institut de Biologie Structurale, UMR 5075, CEA, CNRS, Université Grenoble Alpes, Grenoble, France*

Epstein-Barr virus (EBV) is one of the most widespread viruses in the world; more than 90% of the planet's adult population is infected. Symptomatic primary infection by this *Herpesviridae* corresponds to infectious mononucleosis (IM), which is generally a benign disease. While virus persistence is often asymptomatic, it is responsible for 1.5% of cancers worldwide, mainly B cell lymphomas and carcinomas. EBV may also be associated with autoimmune and/or inflammatory diseases. However, no effective treatment or anti-EBV vaccine is currently available. Knowledge of the proteins and mechanisms involved in the different steps of the viral cycle is essential to the development of effective vaccines. The present review describes the main actors in the entry of the virus into B cells and epithelial cells, which are targets of interest in the development of prophylactic vaccines aimed at preventing viral infection. This review also summarizes the first vaccinal approaches tested in humans, all of which are based on the gp350/220 glycoprotein; while they have reduced the risk of IM, they have yet to prevent EBV infection. The main proteins involved in the EBV latency cycle and some of the proteins involved in the lytic cycle have essential roles in the oncogenesis of EBV. For that reason, these proteins are of interest for the development of therapeutic vaccines of which the objective is the stimulation of T cell immunity against EBV-associated cancers. New strategies aimed at broadening the antigenic spectrum, are currently being studied and will contribute to the targeting of the essential steps of the viral cycle, the objective being to prevent or treat the diseases associated with EBV.

Keywords: Epstein-Barr virus (human gammaherpesvirus 4), EBV-associated cancer, glycoproteins, infectious mononucleosis, latency, lytic cycle, prophylactic vaccine, therapeutic vaccine

OPEN ACCESS

Edited by:
Jonatas Abrahao, Federal University of Minas Gerais, Brazil

Reviewed by:
Graciela Kunrath Lima, Federal University of Minas Gerais, Brazil
Marco Antonio Campos, René Rachou Institute, Oswaldo Cruz Foundation (Fiocruz), Brazil

****Correspondence:***
Raphaële Germi
rgermi@chu-grenoble.fr

Specialty section:
This article was submitted to Virology, a section of the journal Frontiers in Microbiology

Received: *28 April 2021*
Accepted: *20 May 2021*
Published: *22 June 2021*

Citation:
Jean-Pierre V, Lupo J, Buisson M, Morand P and Germi R (2021) Main Targets of Interest for the Development of a Prophylactic or Therapeutic Epstein-Barr Virus Vaccine. Front. Microbiol. 12:701611. doi: 10.3389/fmicb.2021.701611

Abbreviations: Ad5, type 5 adenovirus; CR2, complement receptor type 2; CTLD, C-type lectin domain; DC, dendritic cell; EBER, Epstein-Barr encoded small RNA; EBNA, Epstein-Barr nuclear antigen; EBV, Epstein-Barr virus; EphA2, ephrin receptor tyrosine kinase A2; FBN, fibronectin region; g, gp, glycoprotein; GLA/SE, glucopyranosyl lipid A-stable emulsion; HLA, human leukocyte antigen; IM, infectious mononucleosis; LBD, ligand binding domain; LMP, latent membrane protein; miRNA, microRNAs; MVA, modified vaccinia virus Ankara; NPC, nasopharyngeal carcinoma; NRP1, neuropilin 1; PTLD, post-transplant lymphoproliferative disease; VCA, viral capsid antigen; VLP, virus-like particle; VV, vaccinia virus.

HIGHLIGHTS

- EBV is an oncogenic virus with tropisms for B cells and epithelial cells.
- gp350 vaccines reduce the risk of IM without preventing EBV infection.
- gp350/220, gH, gL, gp42, and gB glycoproteins are targets for neutralizing antibodies.
- Latency proteins are involved in EBV oncogenesis.
- Some lytic cycle proteins may contribute to EBV oncogenesis.
- New vaccine candidates combine latency and lytic cycle antigens.

INTRODUCTION

Epstein-Barr virus (EBV), also known as Human gammaherpesvirus 4, belongs to the *Herpesviridae* family and the *Gammaherpesvirinae* sub-family. This enveloped double-stranded DNA virus with a diameter of 150–200 nm, was discovered in 1964; it initially arose from Burkitt lymphoma tumor cells (Epstein et al., 1964; Mui et al., 2019).

Interhuman transmission of the virus is essentially salivary. In developing countries, almost all children are infected before four; in developed countries, less than 50% of children between five and ten are EBV-seropositive (Niederman and Evans, 1997). While primary EBV infection is more often asymptomatic, it can in some cases be associated with infectious mononucleosis (IM), which is characterized by angina, lymphadenopathy, fever, and fatigue. The later the occurrence of primary infection in the life of an individual (during adolescence or adulthood), the greater the risk of developing IM (Cohen, 2015). Though it is generally a benign disease, in 1% of cases it can entail complications that may be serious (encephalitis, myocarditis, hepatic complications. . .) or disabling; in 10% of patients, chronic fatigue lasts up until 6 months after viral infection (Dasari et al., 2017).

Subsequent to primary infection, EBV persists for a lifetime in the memory B lymphocytes of the infected host, albeit generally without pathological consequences on the individual. However, viral persistence can be associated with the development of cancer. More precisely, EBV is classified in group 1 of human carcinogens. It is the first human oncogenic virus to have been discovered and to this day, it remains the only human pathogen that can immortalize and transform cells *in vitro* (Niedobitek, 1999). Given its B cell tropism, EBV infection can be associated not only with B cell lymphomas (Hodgkin, Burkitt lymphoma), but also with the lymphoproliferative disorders observed in a context of immunodepression, particularly in solid organ transplant patients or in hematopoietic stem cell recipients during the first year following transplantation (Shannon-Lowe et al., 2017). Moreover, given its epithelial cell tropism, EBV is the source of nasopharyngeal cancer and can be associated with gastric carcinomas (Dasari et al., 2019). However, as the virus does not persist in epithelial cells, epithelial tumors arise from the latent viral reservoir present in B cells (Brooks et al., 2016). An estimated 200,000 new cases of EBV-induced cancers occur every year, representing 1.5% of the cancers reported worldwide (Holmes, 2014), and they are responsible for approximately 164,000 deaths a year (Khan et al., 2020).

Epstein-Barr virus may also be associated with the development of inflammatory and autoimmune diseases such as multiple sclerosis (Levin et al., 2010), systemic lupus erythematosus (Ascherio and Munger, 2015), rheumatoid arthritis (Balandraud and Roudier, 2018), and the Sjögren syndrome (Sorgato et al., 2020), but these associations remain controversial.

As it infects over 90% of the world population, EBV is one of the most widespread viruses throughout the planet; however, to this day, no effective treatment or anti-EBV vaccine are available (Pei et al., 2020).

A prophylactic vaccine, which would prevent EBV infection and thereby provide protection against associated diseases, or a therapeutic vaccine, which would directly treat the pathologies associated with EBV, would undeniably be of considerable interest for public health.

The oncogenic potential of EBV precludes its being used in vaccinal projects in an attenuated or inactivated form. That is one reason why development of an anti-EBV vaccine presupposes optimal knowledge of the different elements contributing to the virus's life cycle, namely the viral and cellular proteins implicated in the entry of the virus into host cells (prophylactic vaccines), and the proteins involved in viral persistence (therapeutic vaccines).

VIRUS ENTRY AND PROPHYLACTIC VACCINES

During primary infection, the virus crosses the mucosal epithelial cell barrier by transcytosis and then infects the B cells in the submucosal secondary lymphoid tissues (Tangye et al., 2017).

Epstein-Barr virus entry proceeds in two successive steps, the first of which consists in the tethering of the virus to target cells; viral glycoproteins and cell-adhesion receptors bring nearer and concentrate the virus at the target cell surfaces without triggering cell fusion mechanisms. Contrarily to most of the other enveloped viruses, which utilize only one or two glycoproteins, several viral envelope glycoproteins play a key role in the mechanism of EBV entry into B cells and epithelial cells. The second step consists in the fusion of the virus with the endocytic membrane of the B cells or the plasma membrane of the epithelial cells, a process involving proteins of the fusion machinery highly conserved within the *Herpesviridae* family and cellular entry receptors (Connolly et al., 2011).

Interestingly, the mechanisms of virus entry differ between B cells and epithelial cells. While the entry of EBV into B cells is carried out by endocytosis, EBV enters the epithelial cells by direct fusion of the viral and cellular membranes (Miller and Hutt-Fletcher, 1992).

The Entry of EBV Into B Cells

Glycoprotein 350/220 (gp350/220), expressed on the EBV envelope, is presented in two forms, one in 350 kDa and the other in approximately 220 kDa (alternative splicing of the same

primary RNA transcript). Gp350/220 is strongly implicated in the infection of B cells while its involvement is limited in the infection of epithelial cells.

The entry of EBV into B cells begins with the tethering of gp350/220 via its N-terminal residues 1–470 to the complement receptor type 2 (CR2), also termed CD21, or to the complement receptor type 1 (CD35 or CR21) (Connolly et al., 2011). **Figure 1** outlines in detail the relevant interactions.

Following which, glycoproteins H (gH), L (gL), and 42 (gp42) come into play, regulating the activation of glycoprotein gB, which plays a direct role in membrane fusion (Heldwein, 2016). Taken together, these four glycoproteins are involved in the fusogenic mechanism enabling EBV to enter into B cells.

The gH glycoprotein consists in four major domains ranging from D-I (N-terminal) to D-IV (C-terminal). The gL glycoprotein interacts with the D-I domain to form the heterodimeric gH/gL complex (also termed gp85/gp25) (Matsuura et al., 2010). The gp42 N-terminal part is wrapped around the D-II, D-III, and D-IV domains of gH, and tethered to the KGD-binding motif (present in the D-II domain) (Möhl et al., 2019).

Following the tethering of gp350/220 to CR2, gp42 C-type lectin domain (CTLD situated in the C-terminal) interacts specifically with the β chains of the human leukocyte antigen (HLA) class II molecules (Kirschner et al., 2009; Sathiyamoorthy et al., 2014). The interaction enlarges the hydrophobic pocket of the gp42 CTLD, thereby enabling the gH/gL complex [*via* the glutamine (Q) 54 and lysine (K) 94 residues of the glycoprotein gL] to be tethered to the pre-fusion form of gB (gB residues 450–800) (Mohl et al., 2016). The binding of gH/gL to gB triggers the conformational change of gB into an intermediate form, enabling the insertion of gB fusion loops (hydrophobic residues $WY_{112-113}$ and $WLIW_{193-196}$) into the B cell membrane. Fusion of the viral and cellular bilayers leads to the nucleocapsid release into the cytoplasm of the target cells, as a result of which, gB takes on a stable, post-fusion conformation (Möhl et al., 2017). As of now, only the post-fusion gB structure is known.

The Entry of EBV Into Epithelial Cells

The CR2, to which gp350/220 binds for entry into B cells, is not constantly expressed in epithelial cells. This makes the role of gp350/220 minor in EBV entry into these epithelial cells.

However, the binding of the virus to CR2-negative epithelial cells is five times weaker than its binding to CR2-positive epithelial cells (Chen and Longnecker, 2019).

While interaction between the cellular integrin αVβ1 and the RGD integrin binding motif (arginine/glycine/aspartic acid) of the viral protein BMRF2 is a first step toward facilitated entry, it does not necessarily take place (**Figure 2**; Connolly et al., 2011). The BMRF2 protein can also form a complex with the viral protein BDLF2 (not represented in **Figure 2**). BDLF2 is a type II membrane protein that participates in the rearrangement of the cellular actin network, thereby increasing contacts between the different cells and favoring circulation of the virus from one cell to another (Gore and Hutt-Fletcher, 2009).

Due to the KGD motif of gH, the gH/gL complex of the virus is subsequently tethered to the cellular integrins αVβ5, αVβ6, and αVβ8 (Chesnokova and Hutt-Fletcher, 2011; Connolly et al., 2011). Indeed, the KGD motif is capable of being bound competitively to either the epithelial cellular integrins or to gp42, and the double functionality of the KGD motif of gH is what renders gp42 a major actor in the cellular tropism of EBV (Chen et al., 2014). More precisely, the presence of gp42 masks the KGD motif of gH to epithelial cell integrins and inhibits the fusion of the virus with these cells.

The following step involves the cellular protein EphA2 (ephrin receptor tyrosine kinase A2). The ectodomain of this protein consists in four regions: an LBD region (ligand binding domain) that can be bound to gH/gL and gB, a region rich in cysteine and two fibronectin regions (FBN) that interact with gB. This step is essential, given the fact that in the absence of EphA2, fusion between viral and cellular membranes is reduced by more than 90% and infection by 85% (Chen et al., 2018). Conversely, its overexpression facilitates fusion and infection of the epithelial cells (Chen and Longnecker, 2019).

The fusion process previously described with regard to B cells can consequently be initiated due to interaction between gH/gL and the pre-fusion form of gB, thereby enabling the entry of the virus into epithelial cells.

Another cellular actor has been characterized as having a role in the fusion of EBV with the epithelial cells. Neuropilin 1 (NRP1) is a cellular protease of which the overexpression significantly increases the infection of epithelial cells by EBV; when inhibited, on the other hand, it reduces infection by 50% (Wang et al., 2015). Of note, NRP1 is very weakly expressed by B cells as compared to epithelial cells.

The Tropism of EBV Infection

While gp42 does not play a part in the mechanism of entry of EBV into the epithelial cells, it nevertheless mediates the cellular tropism of the virus (Möhl et al., 2019). More precisely, when neosynthesized virions are exported from B cells, gp42 remains sequestered due to its interaction with HLA class II molecules. As a result, the mature virions exported from B cells are, for the most part, deficient in gp42, and cannot reinfect another B cell (Mohl et al., 2016). Conversely, the absence of gp42 at the surface of the virions promotes infection of the epithelial cells. When the neosynthesized virions are exported from the epithelial cells, which do not contain HLA class II molecules, gp42 is not sequestered and the mature virions are enriched in gp42, a factor facilitating their entrance into the B cells, and so on and so forth (Sathiyamoorthy et al., 2016).

Other EBV Surface Glycoproteins

Other than the just-mentioned glycoproteins, which actively participate in the entrance of the virus into host cells, there exist other glycoproteins that are probably less relevant as possible targets in a vaccine project, and of which the exact roles in the viral cycle remain to be clarified (Johannsen et al., 2004; **Figure 3**).

Among them, we may cite glycoprotein N (gN), which is encoded by the BLRF1 gene, and glycoprotein M (gM), which is composed of several transmembrane domains and is encoded by the BBRF3 gene. Highly conserved in the *Herpesviridae* family,

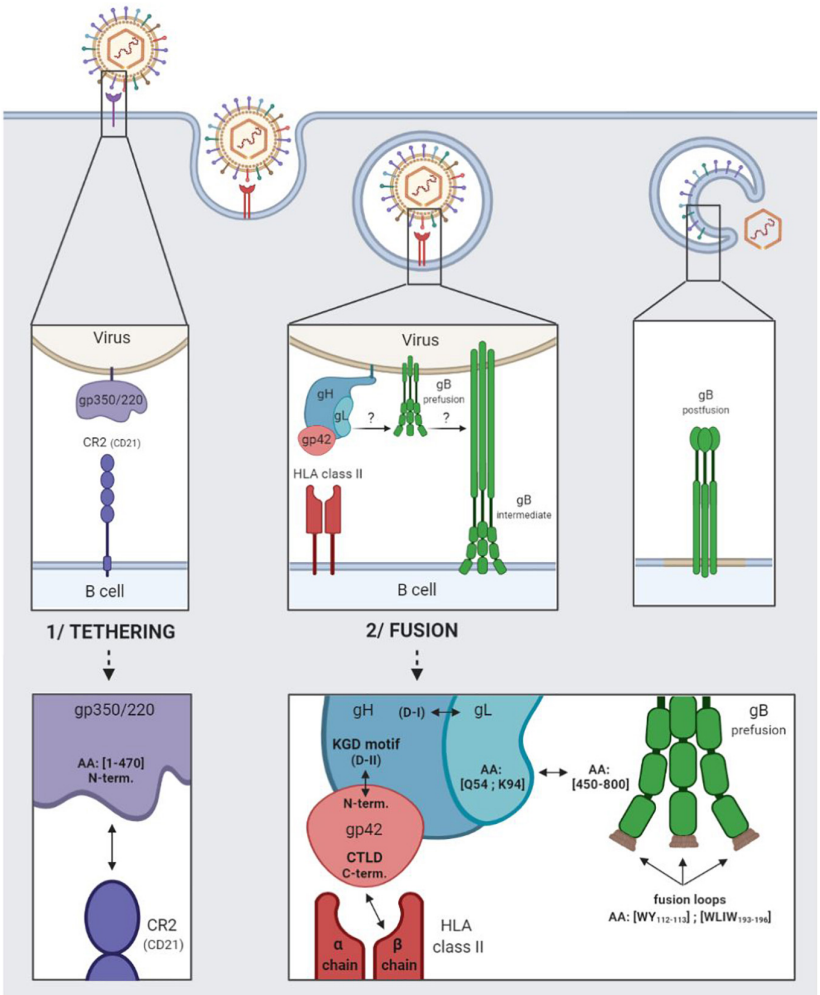


FIGURE 1 | EBV entry into B cells by endocytosis. 1/Tethering of the virus to the B cell involving the viral gp350/220 glycoprotein and the cellular CR2 receptor. 2/Fusion of the viral and cellular membranes involving the viral gH/gL, gp42, and gB glycoproteins and the cellular HLA class II.

these glycoproteins form the gN/gM complex (Lake and Hutt-Fletcher, 2000). Non-expression of the gN/gM complex reduces the binding of the virus to target cells, blocks dissociation of the viral capsid from the cellular membranes (thereby preventing its migration toward the nucleus), and limits the release of new virions (Lake and Hutt-Fletcher, 2000).

Encoded by the BDLF3 gene, glycoprotein 150 (gp150) is capable of binding to the heparan sulfate present at the surface of the epithelial cells, binding that does not provoke the infection of target cells (Chesnokova et al., 2016).

Encoded by the BILF gene, glycoprotein 78/55 (gp78/55) does not seem to have a role in virus entrance (Mackett et al., 1990).

The Main Anti-gp350/220 Prophylactic Vaccines

The initial and principal approaches to prophylactic vaccination naturally targeted gp350/220, which is the most abundant glycoprotein on the surface of EBV and infected cells and represents the main target of anti-EBV neutralizing antibodies (Cohen et al., 2011). The different vaccinal approaches are summarized in **Figure 4**.

The first anti-EBV vaccine clinical trial was developed in 1995 by a Chinese team that used as its vector, a live vaccinia virus (VV) expressing gp350/220. The vaccine was administered in a single dose to: (i) EBV and VV-seropositive adults; (ii) EBV-seropositive juveniles who were nonetheless VV-seronegative; (iii) children who were seronegative for the two viruses. The outcomes were analyzed 16 months after vaccine administration. In adults, due to their anti-VV immunity, the VV did not multiply and the vaccine had no effect. On the contrary, in vaccinated juveniles, anti-EBV neutralizing antibody titers increased, and among the nine seronegative vaccinated children, only three were EBV-infected, whereas all 10 of the non-vaccinated children were (Gu et al., 1995). Due to the biosafety standards imposed in vaccinology, this initial clinical study was not pursued.

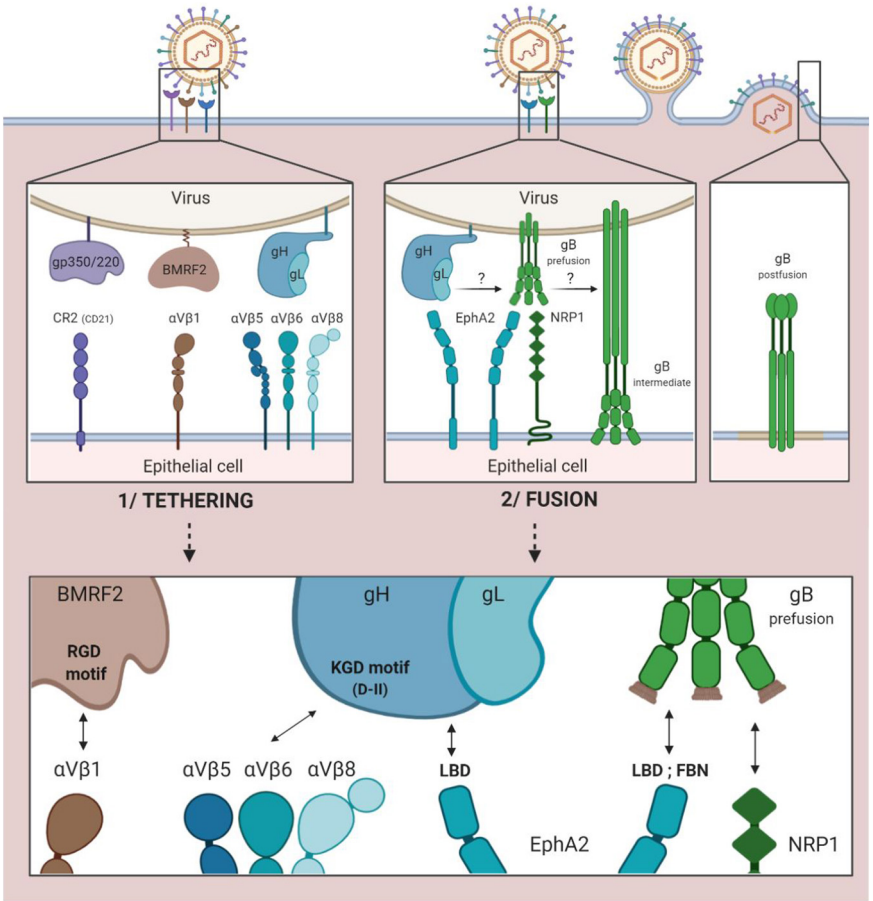


FIGURE 2 | EBV entry into epithelial cells by direct fusion. 1/Tethering of the virus to the epithelial cell involving the viral gp350/220, BMRF2 and gH/gL glycoproteins and the cellular CR2 receptor (case of CR2+ epithelial cells) and cellular integrins (αVβ1, 5, 6, and 8). 2/Fusion of viral and cellular membranes involving the viral gH/gL and gB glycoproteins and the cellular EphA2 and NRP1 proteins.

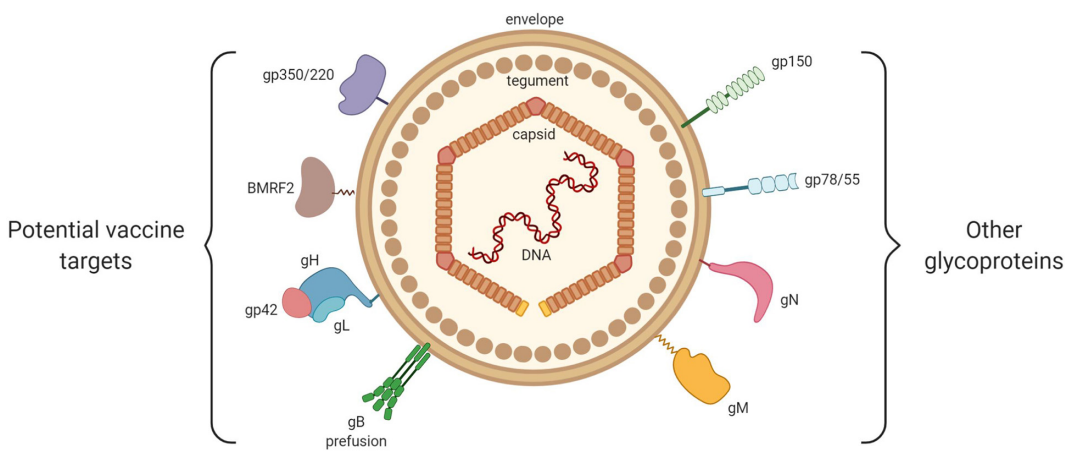


FIGURE 3 | General structure of EBV. From inside to outside are represented: the EBV double-stranded DNA, the icosahedral capsid, the tegument and the envelope. On the left are shown the viral envelope glycoproteins that play a key role in EBV entry into host cells (gp350/220, BMRF2, gH/gL, gp42, and gB). On the right are shown the viral glycoproteins with little or no involvement in EBV entry (gp150, gp78/55, gN, and gM).

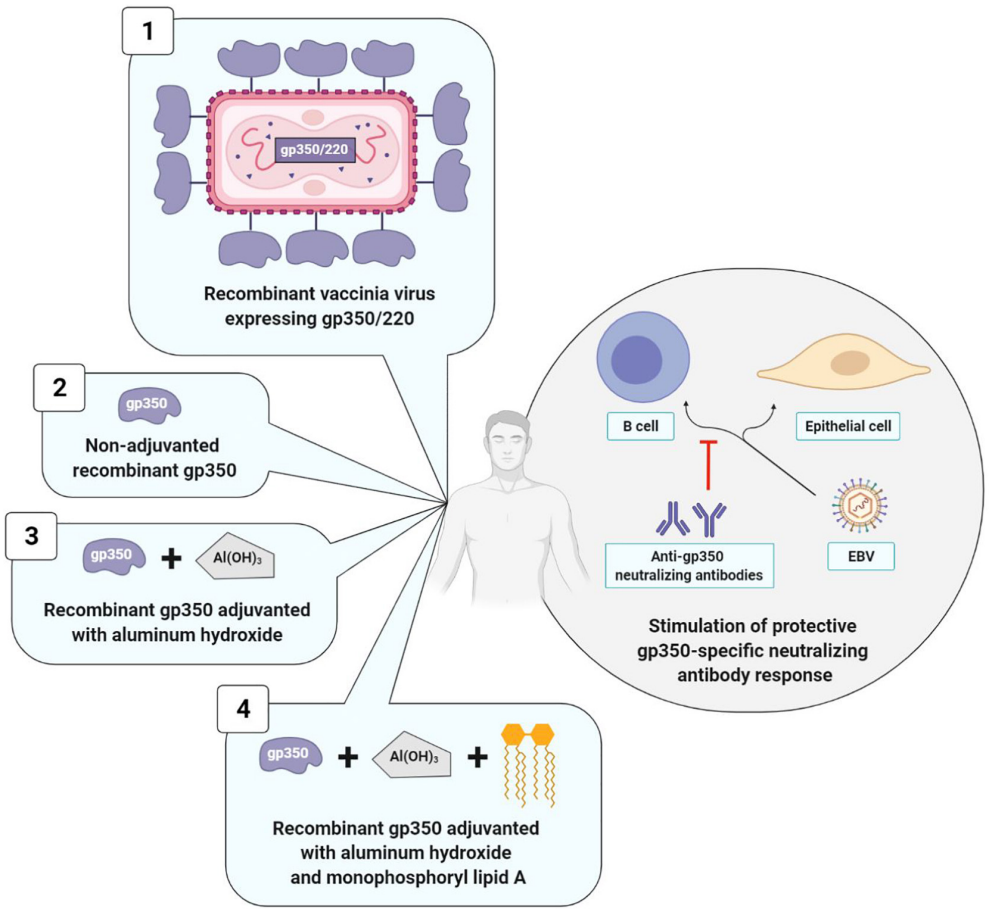


FIGURE 4 | Prophylactic vaccine strategies using gp350/220 as antigen. The four types of strategies elicit a neutralizing anti-gp350 immune response. Strategies with adjuvant (3 and 4) induce higher anti-gp350 antibody titers than strategy without adjuvant (2). Strategy with monophosphoryl lipid A (4) induce a higher and long-lasting neutralizing anti-gp350 humoral response than strategy without monophosphoryl lipid A (3). Data collected for strategy 1 are not comparable to those of other strategies. 1, Gu et al., 1995; 2, Moutschen et al., 2007; 3, Moutschen et al., 2007; Rees et al., 2009); 4, Moutschen et al., 2007; Sokal et al., 2007.

It was only in 2007 that Moutschen et al. (2007) published the results of two new clinical trials (one phase I, and the other phase I/II), which were carried out with 148 healthy adults. Both of these studies demonstrated the safety and immunogenicity of a soluble gp350 recombinant monomer, associated or not with an adjuvant. This candidate vaccine was obtained from CHO (Chinese Hamster Ovary) cells expressing the gp350/220 gene having undergone splice site mutation, thereby preventing formation of the gp220 isoform. The preparations associating an adjuvant (aluminum hydroxide alone or combined with monophosphoryl lipid A) induced anti-gp350 antibody titers significantly higher than preparations without an adjuvant and the three tested candidate vaccines were well-tolerated (Moutschen et al., 2007).

In order to confirm the safety and immunogenicity of this anti-EBV vaccine and to assess its efficacy, Sokal et al. (2007) carried out a phase II, double blind, randomized trial. The soluble gp350 recombinant monomer associated with aluminum hydroxide/monophosphoryl lipid A was distributed in three doses to 181 EBV-seronegative young adults and compared to a placebo consisting in aluminum hydroxide alone. A durably neutralizing anti-gp350 antibody response was observed over a period exceeding 18 months in 98.7% of the vaccinated patients. While this humoral response did not suffice to prevent EBV infection, it significantly reduced (78%) IM incidence in the vaccinated group; IM occurrence was 4.8 times higher in the placebo group than in the vaccinated group (Sokal et al., 2007). In conclusion, the anti-gp350 neutralizing antibodies seemed to attenuate the severity of the disease associated with EBV infection; on the other hand, they did not prevent the infection itself (Tangye et al., 2017).

Two years later, another phase I study assessed the effect of the gp350 recombinant monomer associated with aluminum hydroxide in the prevention of post-transplant lymphoproliferative disease (PTLD) in 16 EBV-seronegative children with chronic renal disease who were awaiting transplantation. The objective of this vaccine was to stimulate anti-EBV humoral immunity prior to transplantation in view of reducing PTLD appearance and severity. The two tested doses of the vaccine (12.5 and 25 μg) were well-tolerated and

induced similar anti-gp350 antibody titers in all of the vaccinated patients; however, only four patients developed anti-gp350 neutralizing antibodies. What is more, anti-gp350 immune responses (neutralizing or not) rapidly declined a few weeks after the last injection, and 26 weeks after transplantation, the EBV viral loads measured in the blood by PCR were similar between vaccinated and non-vaccinated children. A preventive effect of the vaccine on PTLD consequently appears unlikely (Rees et al., 2009).

VIRAL LATENCY, VIRAL REPLICATION, AND THERAPEUTIC VACCINES

The second anti-EBV vaccine strategy consists in stimulating T cell immunity to EBV and/or restoring immune control in patients with EBV-associated cancers. So-called therapeutic vaccines rely primarily on the latency proteins implicated in the transformation and immortalization of infected cells and, consequently, the development of cancers. Some proteins in the lytic cycle also play a major role in the establishment of viral persistence. Therapeutic vaccines could represent an alternative to the radiotherapies and chemotherapies usually applied in treatment of EBV-associated cancers, which are responsible for considerable adverse effects.

Development of therapeutic vaccines necessitates knowledge of several different steps and of the proteins involved in the latency cycle and the EBV replication cycle in infected individuals.

The EBV Latency Cycle

After primary infection, EBV persists as an episomal latent form in the infected memory B lymphocytes. At this stage, there exists a balance between latency, replication of the virus in the organism and elimination of the infected cells by immune responses. Latency is characterized by the expression of viral latency genes, six of which code for nuclear proteins EBNA (Epstein-Barr nuclear antigen), three for membrane proteins LMP (latent membrane protein) and two for non-coding RNAs EBER (Epstein-Barr encoded small RNA). A large number of viral-derived micro-RNAs (miRNA) are likewise expressed (44 stemming from the BART gene and three from the BHRF1 gene) (Iizasa et al., 2020; **Figure 5**). These different latency proteins and non-coding RNAs of EBV contribute to the capacity of the virus to immortalize and ensure B cell proliferation *ad infinitum*. There exist different expression profiles for latency genes, enabling them to persist in B lymphocytes during cell division. These different profiles are expressed during the different natural phases of latency establishment but also in different EBV-associated cancer pathologies. In the three types of latency, EBNA1 expression allows the long-term persistence of the EBV genome as an episome and its replication by cellular DNA polymerase. During the latency phase, the virus is replicated synchronously with the infected memory B cells, which thereby become the main site of EBV persistence (Andrei et al., 2019).

Type III latency corresponds to the expression in naive B cells of all the different latency genes. This type of latency is found mainly in non-Hodgkin lymphomas such as PTLD or in EBV-associated diffuse large B cell lymphomas. In most cases, this type of lymphoma appears in immunocompromised persons in whom reduced immunosurveillance entails uncontrolled B cell proliferation, which fosters the appearance of genetic alterations and the development of cancer (Cohen, 2015).

Type II latency corresponds to the expression, in germinal centers, of EBNA1, LMP1, LMP2, non-coding EBER RNAs and miRNA and is found in EBV-associated Hodgkin lymphomas and nasopharyngeal carcinomas (NPC).

Latency I/0 is present in memory B cells, in which only EBNA1 and non-coding EBER RNAs and miRNA are expressed and is found in Burkitt lymphomas and EBV-associated gastric carcinomas.

In Hodgkin and Burkitt lymphomas and NPC, EBV expression is consequently limited to a few latency genes, and reduced host immune response is of limited importance. On the other hand, the role of cellular gene mutations is essential in the pathogenesis and the development of these EBV-associated cancers (**Figure 5**).

As a result, the EBNA1 protein (expressed in all the EBV-associated malignant tumors), the signaling membrane proteins LMP1 (oncogene implicated in cell transformation and survival) and LMP2 (regulator of virus reactivation from latency) as well as the non-coding RNAs (implicated in immune escape rather than cell transformation) represent promising targets for the composition of an anti-EBV therapeutic vaccine (Dasari et al., 2019; Münz, 2019).

The EBV Lytic Cycle

As the study of EBV-associated cancers has for many years been exclusively focused on latency proteins, the role of the lytic cycle seems to have been underestimated.

The EBV lytic cycle (also known as a replicative or multiplication cycle) is divided into three sequential phases (immediate early, early and late) during which more than 80 proteins are expressed.

The first genes to be transcribed during the immediate early phase are the BZLF1 and BRLF1 genes, which code the ZEBRA (or Zta or Z) and the Rta (or R) transcription factors respectively. The ZEBRA and Rta proteins then activate the early gene promoters, which code for either the replication complex (including viral DNA polymerase, its processivity factor, thymidine kinase and the helicase-primase complex) or the proteins involved in late gene expression. After which, late genes are expressed, coding for structural proteins such as viral capsid antigen (VCA), the viral protease necessary for maturation of the capsid and the envelope glycoproteins; taken together, they enable construction of new viral particles (Kenney and Mertz, 2014; **Figure 6**).

The immediate early ZEBRA and Rta proteins are more often the target of TCD8+ cells, while the early and late proteins are more often the target of TCD4+ cells (Cohen, 2018). Thus, EBV surface glycoproteins (late proteins) are not simply the target of humoral immunity with neutralizing antibodies; they are also targeted by glycoprotein-specific CD4+ T cells, which are capable of recognizing newly EBV-infected cells (Brooks et al., 2016).

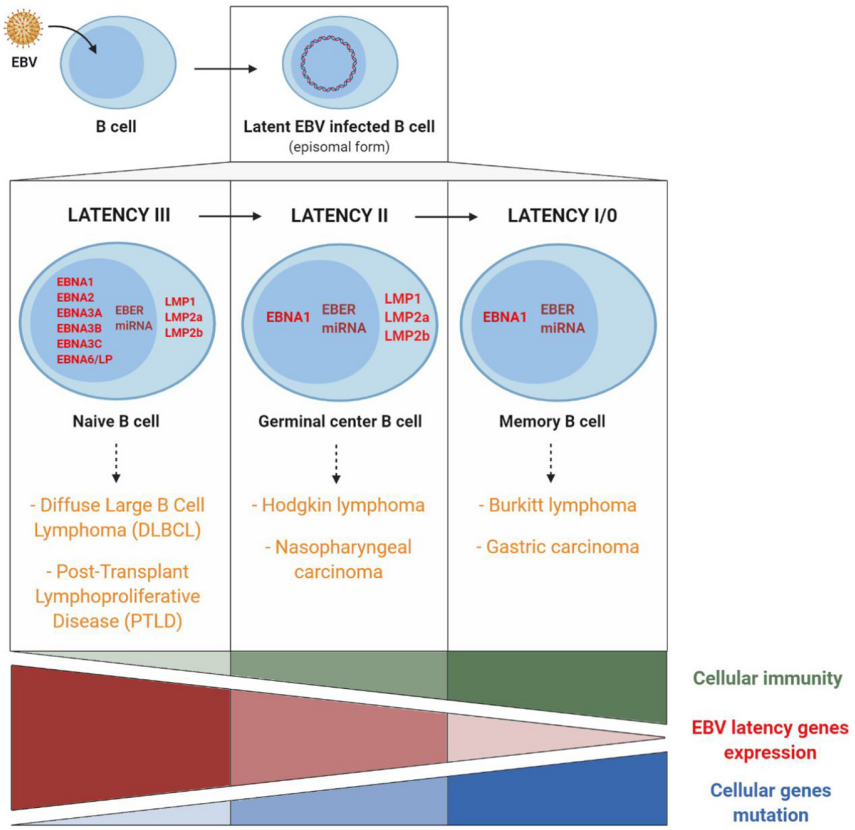


FIGURE 5 | EBV latency gene expression profiles and involvement of cellular immunity, of EBV latency gene expression, and of cellular gene mutations in EBV-associated cancers. Cancers associated with EBV type III latency are mainly found in immunocompromised patients. Conversely, cancers associated with EBV type I latency are mainly found in immunocompetent patients and are related to host cell gene mutations. Adapted from Rickinson and Kieff (2013) and Lupo et al. (2019).

At present, numerous arguments emphasize the role of proteins expressed during the lytic cycle in the initiation and development of EBV-associated cancers. The partial success of anti-EBV immunotherapy targeting the antigens of the lytic cycle in EBV-associated cancers is one of these arguments (Münz, 2020). In addition, it has been shown that humanized mouse models developed fewer B cell lymphomas when they were infected with a deleted EBV mutant of the lytic gene BZLF1 than with a non-muted EBV (Ma et al., 2011).

As a result, lytic cycle proteins – and, more precisely, immediate early proteins – may be of interest in possible association of latency proteins with the conception of a therapeutic vaccine. A vaccine composed of the main lytic proteins could stimulate the TCD4+ and TCD8+ cellular immunity specific to these antigens and eliminate infected cells, thereby limiting the formation of new virions, as well as the infection and transformation of new cells.

The Main Therapeutic Anti-EBV Vaccines Tested in Humans

Most of the therapeutic anti-EBV vaccines have been developed for the treatment of NPC, which represents the third most frequent cause of cancer in southern China (Straathof et al., 2005). NPC is an epithelial tumor closely associated with EBV (100% of cases) in which the tumor cells express a type II latency profile. The EBNA1 and LMP2 proteins significantly contribute to the transformation of normal cells into cancer cells. The EBNA1 protein includes epitopes that are targeted mainly by TCD4+ cells, and to a lesser extent by TCD8+ cells. Conversely, the LMP2 protein contains epitopes that are targeted mainly by TCD8+ cells, and only to a small extent by TCD4+ cells (Dasari et al., 2019).

That is why EBNA1 and LMP2 are the two mainly targeted antigens in therapeutic vaccines.

The three therapeutic vaccine strategies are summarized in **Figure 7**. The first one is based on dendritic cells (DC) and the second on viral vectors, while the third is mixed.

Dendritic cells based strategy is situated at the interface of vaccination and immunotherapy. It was tested during a phase I clinical study involving 16 patients at an advanced stage of NPC. Autologous DC were initially harvested and then pulsed *ex vivo* with different LMP2 epitope peptides (A1101, A2402, or B40011) before being reinjected into the inguinal lymph nodes of patients. DC are antigen presenting cells that are of paramount importance in the induction of T cell immune response. In this Chinese study,

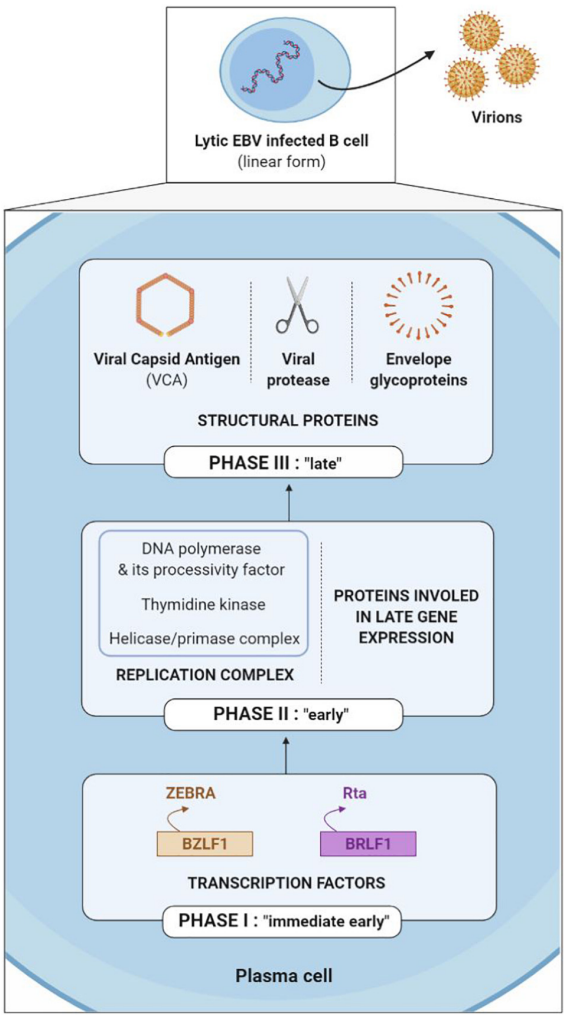


FIGURE 6 | Sequential expression of genes involved in the EBV lytic cycle.

TCD8+ immune responses directed against LMP2 epitopes were observed in nine vaccinated patients, and partial tumor reduction was noted in two of them. However, 6 months after the first injection, the LMP2-specifc TCD8+ cells had reverted to their level prior to vaccination (Lin et al., 2002).

The second vaccine technology uses a viral vector [adenovirus or modified Vaccinia virus Ankara (MVA)]. The genome of these viral vectors is modified by insertion of a DNA sequence of one or more EBV antigens. A phase I clinical study carried out in 24 advanced stage NPC patients demonstrated the safety and immunogenicity of a recombinant type 5 adenovirus (Ad5) (one of the least pathogenic for humans) expressing LMP2 protein (Borovjagin et al., 2014). Surprisingly, the vaccine significantly increased not the level of TCD8+ lymphocytes, but rather the level of TCD4+ lymphocytes alone, and only in patients vaccinated with the highest dose of the tested vaccine (Si et al., 2016).

In their respective approaches to viral vector, two teams have succeeded in improving TCD8+ immune response (Lin et al., 2002) by associating antigen EBNA1 epitopes with LMP2 epitopes and by utilizing the attenuated and recombinant MVA virus as viral vector. In fact, TCD4+ cells assume a crucial role in maintenance of a durably effective memory CD8+ T cell response (Taylor et al., 2004). Known as MVA-EL, the modified virus coded for a fusion protein containing the C-terminal motif of protein EBNA1 as well as the whole LMP2 protein.

An initial phase I clinical trial was carried out in Hong Kong in 2013 including 18 NPC patients who had been in remission for more than 12 weeks, the objective being to assess the safety and immunogenicity of the MVA-EL vaccine (Hui et al., 2013). The same clinical trial took place in the United Kingdom in 2014, involving 14 patients (Taylor et al., 2014). In 15 out of the 18 Hong Kong patients and eight out of the 14 British patients having received three intradermic MVA-EL vaccines 3 weeks apart, the T cell immune response directed against at least one of the two vaccine antigens (EBNA1 or LMP2) was increased.

Conducted in parallel, the two studies showed that the MVA-EL candidate vaccine was well-tolerated and that it increased specific TCD8+ and TCD4+ immune responses.

The third strategy is mixed; while it brings into play the DC, the tumor antigens are incorporated in an adenoviral vector.

A phase II clinical trial carried out in 19 advanced stage NPC patients assessed the safety and the antitumor effects of DC transduced with a recombinant Ad5 vector and coding for a truncated LMP1 and a full-length LMP2 protein. While the vaccine was well-tolerated, its clinical efficacy was limited; no significant increase in LMP1-specific or LMP2-specific T cells was observed *in vivo*, and only three out of the 16 vaccinated patients showed partial clinical benefits (Chia et al., 2012).

OPTIMIZATION OF THE ANTI-EBV VACCINE STRATEGIES

While the first studies on anti-EBV vaccines seem encouraging, improvements with regard to immunogenicity and choice of vaccine epitopes appear necessary, the objective being to envisage a more effective vaccine. New vaccine formulations have consequently been developed.

As concerns prophylactic vaccines, while gp350/220 was the first identified and remains the most widely studied vaccine target, the glycoproteins gH/gL, gp42, and gB have also been identified as neutralizing antibody targets (Sathiyamoorthy et al., 2017; Snijder et al., 2018). In addition, mixed approaches combining EBV lytic cycle and viral cycle proteins are being developed (**Figure 8**).

As concerns therapeutic vaccines, homologous vaccination using the same viral vector for the prime and the boost injections is being superseded by more effective strategies of heterologous vaccination using two distinct vaccine formulations (**Figure 9**).

Optimization of the Prophylactic Vaccines

Anti-gp350/220 Strategies

Improvement of soluble monomeric gp350 vaccines initially consisted in developing gp350/220 tetramers that succeeded in

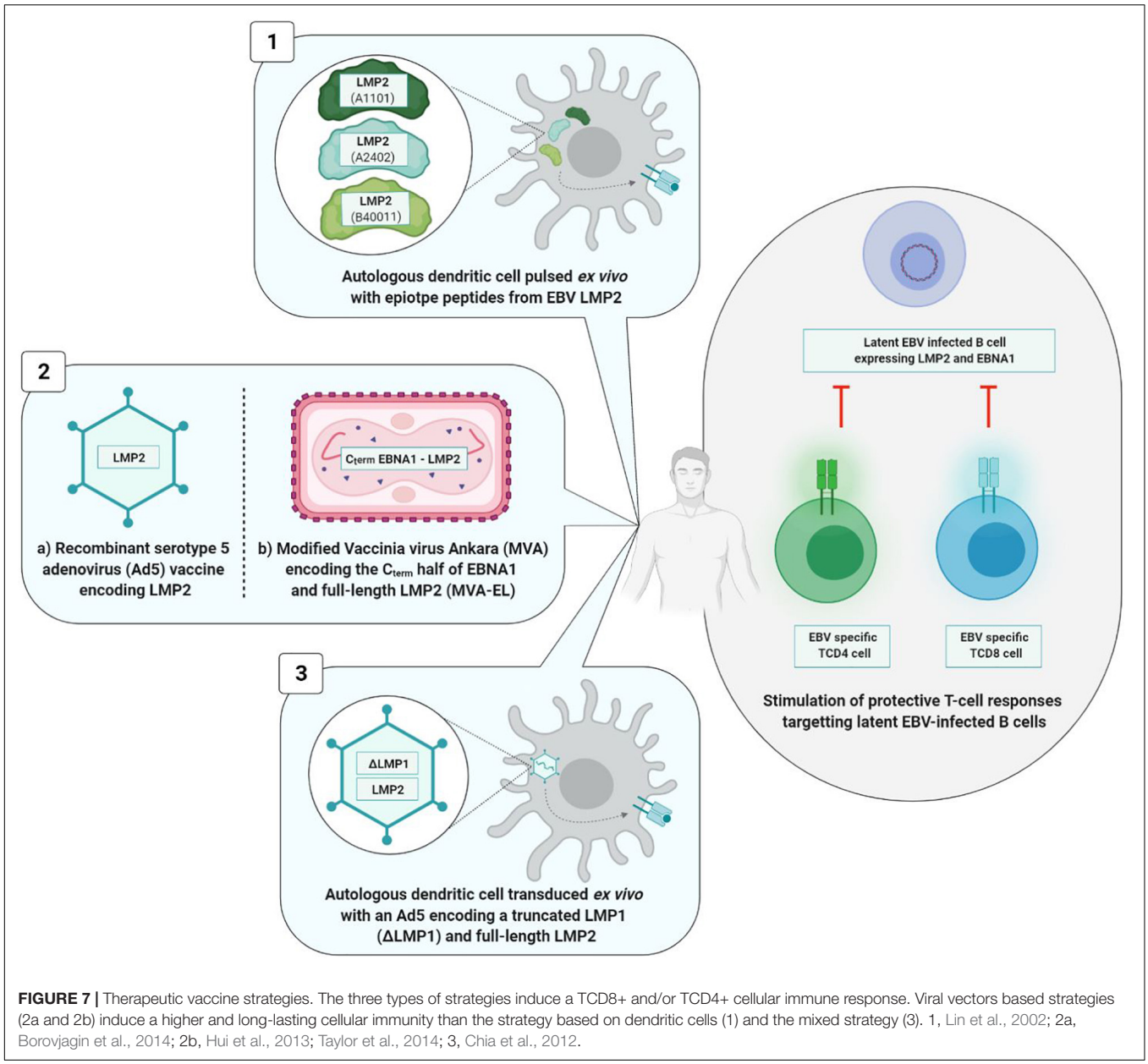


FIGURE 7 | Therapeutic vaccine strategies. The three types of strategies induce a TCD8+ and/or TCD4+ cellular immune response. Viral vectors based strategies (2a and 2b) induce a higher and long-lasting cellular immunity than the strategy based on dendritic cells (1) and the mixed strategy (3). 1, Lin et al., 2002; 2a, Borovjagin et al., 2014; 2b, Hui et al., 2013; Taylor et al., 2014; 3, Chia et al., 2012.

heightening the levels of anti-gp350/220 neutralizing antibodies in BALB/c mice (Cui et al., 2013).

Another approach consisted in developing self-assembling ferritin nanoparticles expressing an epitope of the gp350/220 binding site at cellular CR2. Using this approach, it became possible to multiply by 10 (in the macaque monkey) and by 100 (in the BALB/c mouse) the anti-gp350/220 antibody titers (Kanekiyo et al., 2015).

Another strategy succeeded in multiplying by 10–100 the titers of neutralizing anti-gp350/220 antibodies in BALB/c mice by fusing the gp350/220 ectodomain to the Fc fragment of mouse IgG2a (Zhao et al., 2018).

Another team developed viral particles without viral DNA termed known as VLPs (virus-like particles); arising from nucleocapsid HBc149 of the hepatitis B virus (VLP-HBc149-gp350/220), they constitute a combination of three epitopes derived from the gp350/220-CR2 binding site. In immunized BALB/c mice, VLPs have yielded high levels of neutralizing antibodies (Zhang et al., 2020).

Lastly, Kang et al. (2021) demonstrated the immunogenicity of the first 425 residues of the gp350/220 ectodomain expressed on two different self-assembled nanoparticles that mimic the shape and size of EBV: lumazine synthase (LS) and I3-01, two capsid-forming enzymes. The $gp350/220_{1-425}$-LS and $gp350/220_{1-425}$-I3-01 nanoparticles adjuvanted with aluminum hydroxide or MF59 elicited in mice over 133- and 65-fold higher neutralizing antibody titers, respectively, than that induced by the corresponding gp350 monomer (Kang et al., 2021).

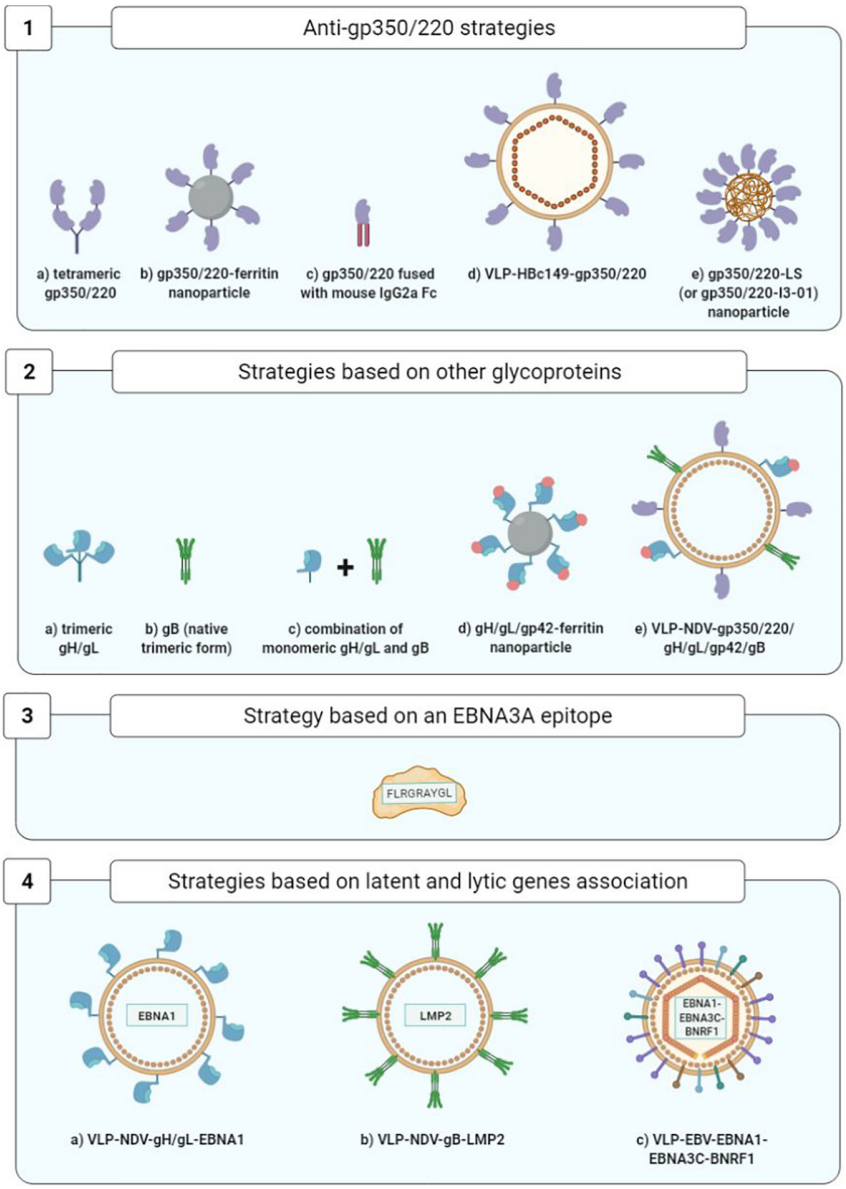


FIGURE 8 | Optimization of prophylactic vaccine strategies. 1a, Cui et al., 2013; 1b, Kanekiyo et al., 2015; 1c, Zhao et al., 2018; 1d, Zhang et al., 2020; 1e, Kang et al., 2021; 2a,2b, Cui et al., 2016; 2c, Cui et al., 2021; 2d, Bu et al., 2019; 2e, Escalante et al., 2020; 3, Elliott et al., 2008; 4a,4b, Perez et al., 2017; 4c, van Zyl et al., 2018.

While these new strategies have yet to be compared with one another, all of them seem to be more immunogenic than soluble gp350 monomers and represent promising attempts to enhance the immunogenicity of anti-gp350/220 vaccines.

Improvement of anti-gp350/220 vaccines requires not only modifications of gp350/220 structure, but also the use of new adjuvants capable of optimizing anti-gp350/220 immune response. This is particularly the case with glucopyranosyl lipid A-stable emulsion (GLA/SE), which has significantly and durably increased the levels of neutralizing antibodies and anti-gp350/220 T cell response in vaccinated mice and rabbits (Heeke et al., 2016).

Strategies Based on the Other EBV Glycoproteins of Interest

The glycoproteins gH/gL and gB assume a preponderant role, not only in EBV entry into B cells, but also (contrarily to gp350/220) into epithelial cells (Snijder et al., 2018).

A study conducted on rabbits immunized with the monomeric or trimeric forms of the gH/gL proteins, the trimeric native form of gB and the monomeric or tetrameric forms of gp350/220 showed that the animals produced higher titers of neutralizing antibodies against the gH/gL and gB glycoproteins than against gp350/220. Moreover, the multimeric forms were more immunogenic than their monomeric counterparts, and whatever

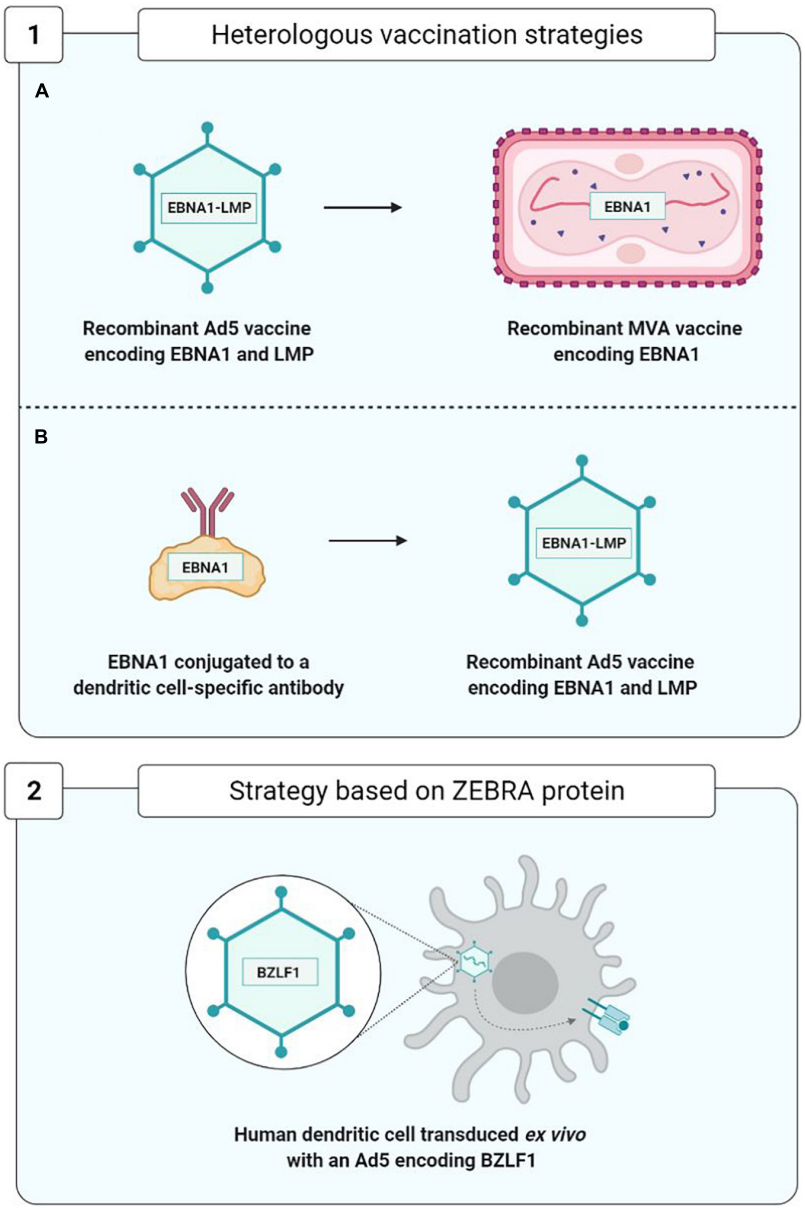


FIGURE 9 | Optimization of therapeutic vaccine strategies. **(1A,1B)**, Rühl et al., 2019; 2, Hartlage et al., 2015.

its form, the gH/gL glycoprotein was the most immunogenic (Cui et al., 2016).

Five years later, the same team showed that immunization of rabbits with the combination of gH/gL (monomeric form) and gB (trimeric native form) elicited higher neutralizing antibody titers (for both B cells and epithelial cells) than that induced by gH/gL or gB alone. In addition, they demonstrated that sera from rabbits immunized with this combination of gH/gL and gB decreased the EBV load in peripherical blood of humanized mice and protected these mice from death caused by lethal dose EBV challenge (Cui et al., 2021).

Another team compared the humoral responses developed by macaques against ferritin nanoparticles expressing gH/gL glycoproteins alone or combined with gp42, to nanoparticles expressing gp350/220. The gH/gL/gp42-ferritin vaccine complex was found to confer a neutralizing antibody titer highly superior to that of the gp350-ferritin complex. The researchers also noted that while the addition of gp42 to the gH/gL nanoparticles multiplied by four to eight times the titers of the antibodies that neutralize B cell infection, it had no significant impact on the epithelial cells. Finally, association of the gp350-ferritin nanoparticles and gH/gL/gp42-ferritin nanoparticles yielded higher neutralization levels than with gH/gL/gp42-ferritin alone (Bu et al., 2019).

To conclude, a recent study evaluated the injection in rabbits of Newcastle disease virus-like particles (NDV-VLP)

expressing the five viral glycoproteins essential to the entry of EBV in target cells (gp350/220, gp42, gH, gL, and gB) in their association with adjuvants (aluminum hydroxide combined with monophosphoryl lipid A). This pentavalent (5-in-1) vaccine stimulated production of antibodies specific to the five glycoproteins capable of neutralizing EBV infection of the B cells as well as the epithelial cells. The IgG anti-gp350/220 and anti-gB levels were markedly higher than the IgG anti-gp42 and anti-gH/gL levels (Escalante et al., 2020).

Strategy Based on the EBNA3A Latency Gene

The objective of one original approach toward an anti-EBV prophylactic vaccine has been to prevent IM development stimulating TCD8+ immune cell response directed against EBNA3A and capable of controlling the expansion of EBV-infected B cells. Given the transformative capacity of EBNA3A, an epitope of this viral antigen (FLRGRAYGL) has been used and associated with the tetanic anatoxin as an adjuvant, in an oil-in-water emulsion. The safety and immunogenicity of this vaccine epitope were demonstrated during a phase I clinical trial bringing together 14 healthy adult volunteers, all of whom were EBV-seronegative. They were monitored for up to 12 years after the vaccination, and it was shown that four out of the eight persons having received a low dose of the vaccine (5 μg) were EBV-infected but did not develop IM. As for the two persons vaccinated with a high dose of vaccine epitope (50 μg), one was infected with EBV and developed minimal IM symptoms. Lastly, two out of the four persons having received a placebo were infected with EBV, and one of them developed IM. However, the small study population precludes definitive conclusions on the efficacy of this candidate vaccine in IM prevention (Elliott et al., 2008).

Strategies Based on the Association of Latency Genes and the Lytic Cycle at the VLP Surface

In the BALB/c mice, Perez et al. (2017) studied the vaccine combinations gH/gL-EBNA1 and gB-LMP2 expressed by NDV-VLP and compared them to NDV-VLP-gp350/220. The mice immunized with NDV-VLP-gH/gL-EBNA1 or NDV-VLP-gB-LMP2 developed an EBNA1 and LMP2-specific T cell response. The neutralizing antibody titers of these mice were higher than those of the mice immunized with NDV-VLP-gp350/220 (Perez et al., 2017).

Another, even more audacious strategy consisted in utilizing VLPs directly derived from non-infectious and non-oncogenic EBV particles (without viral DNA), modified to express the latency proteins EBNA1 and EBNA3C fused with protein BNRF1 (the main EBV tegument protein) (Pavlova et al., 2013). Injection of these modified EBV-VLPs in humanized mice proved conducive to development of a TCD4+ immune cell response that was specific to both structure proteins (BNRF1) and latency proteins (EBNA1 and EBNA3C) (van Zyl et al., 2018, 2019).

Optimization of the Therapeutic Vaccines

In order to improve the immunogenicity of anti-EBV therapeutic vaccines based on viral vectors, new approaches to heterologous vaccination have appeared (**Figure 9**). They consist, during the boost injection, in utilizing a construct different from that of the prime injection. This approach limits the appearance of anti-vector neutralizing antibodies and enhances vaccine efficacy; the two trials carried out by Rühl et al. (2019) constitute a good example.

In the first trial, in the prime injection the team used a recombinant Ad5 viral vector expressing an EBNA1-LMP polyepitope, and in the boost injection, they employed a recombinant and attenuated MVA viral vector expressing the EBNA1 protein.

In another trial, in the prime injection they used a monoclonal antibody protein construct directed against the DEC-205 receptor of the DC attached to protein EBNA1. This construct facilitates the entry and presentation of protein EBNA1 by means of the DCs. The prime injection was boosted by the injection of an Ad5 viral vector expressing the EBNA1-LMP polyepitope.

These two approaches have enabled development of a TCD4+ and TCD8+ EBNA1-specific cell response effectively protecting the mouse from the T and B cell lymphomas expressing EBNA1, thereby justifying their use as therapeutic anti-EBV vaccines in future clinical studies (Rühl et al., 2019).

Another therapeutic vaccination strategy, based on the capacity of the ZEBRA protein to initiate a transition from the viral latency phase to the lytic phase, was used on a hu-PBL-SCID model of mice capable of developing EBV-associated lymphoproliferative diseases (Tang et al., 2016). This vaccine strategy employed human DC transduced with a recombinant Ad5 encoding the ZEBRA protein and enabling the development of ZEBRA-specific TCD8+ cell responses capable of recognizing and eliminating EBV-transformed cells and of significantly delaying the death of mice suffering from lymphoproliferative diseases (Hartlage et al., 2015).

DISCUSSION

Notwithstanding all the efforts expended, up until now no commercialized anti-EBV vaccine able to prevent infection, IM or cancers associated with EBV has been developed. Moreover, anti-EBV vaccination continues to encounter major obstacles (Balfour, 2014).

As concerns prophylactic vaccines, the objective of sterilizing immunity completely preventing the infection appears difficult to achieve in humans. Indeed, it has been shown than an individual can host multiple EBV strains probably acquired by multiple infections. To put it another way, anti-EBV immunity does not prevent reinfections (Walling et al., 2003). The risk of a vaccine transitorily preventing infection would be to put off primary EBV infection until an age when IM infection is more frequent (Dunmire et al., 2018).

If the objective of an anti-EBV vaccine is to limit occurrence of the cancers associated with the virus, the low incidence of these cancers and the sizable time they take to appear considerably complicates clinical trials capable of proving vaccinal efficacy.

Last but not least, another major obstacle consists in the absence of efficient animal models, affecting not only

the development of prophylactic vaccines, but also research on therapeutic vaccines (Rühl et al., 2020).

Present-day hopes for an anti-EBV vaccine reside in a combination of several strategies and on the use of several latency and lytic cycle proteins that could induce a broad spectrum of neutralizing antibodies and TCD8+ and TCD4+ cellular response. The objective being to block the essential steps of the viral cycle, from the entry through the persistence of the target cells, and to provide protection against EBV infection (prophylactic effect) and the associated diseases (therapeutic effect).

AUTHOR CONTRIBUTIONS

VJ-P and RG drafted the manuscript. All authors contributed to manuscript revision, read, and approved the submitted version.

ACKNOWLEDGMENTS

We thank Jeffrey Arsham (native English speaker) for the translation of our review.

REFERENCES

Andrei, G., Trompet, E., and Snoeck, R. (2019). Novel therapeutics for Epstein–Barr Virus. *Molecules* 24:997. doi: 10.3390/molecules24050997

Ascherio, A., and Munger, K. L. (2015). "EBV and autoimmunity," in *Epstein Barr Virus Volume 1 Current Topics in Microbiology and Immunology*, ed. C. Münz (Cham: Springer International Publishing), 365–385.

Balandraud, N., and Roudier, J. (2018). Epstein-Barr virus and rheumatoid arthritis. *Joint Bone Spine* 85, 165–170. doi: 10.1016/j.jbspin.2017.04.011

Balfour, H. H. (2014). Progress, prospects, and problems in Epstein-Barr virus vaccine development. *Curr. Opin. Virol.* 6, 1–5. doi: 10.1016/j.coviro.2014.02.005

Borovjagin, A. V., Gomez-Gutierrez, J. G., Shirwan, H., and Matthews, Q. L. (2014). "Adenovirus-based vectors for the development of prophylactic and therapeutic vaccines," in *Novel Technologies for Vaccine Development*, eds I. S. Lukashevich and H. Shirwan (Vienna: Springer Vienna), 203–271.

Brooks, J. M., Long, H. M., Tierney, R. J., Shannon-Lowe, C., Leese, A. M., Fitzpatrick, M., et al. (2016). Early T cell recognition of B cells following Epstein-Barr Virus infection: identifying potential targets for prophylactic vaccination. *PLoS Pathog.* 12:e1005549. doi: 10.1371/journal.ppat.1005549

Bu, W., Joyce, M. G., Nguyen, H., Banh, D. V., Aguilar, F., Tariq, Z., et al. (2019). Immunization with components of the viral fusion apparatus elicits antibodies that neutralize Epstein-Barr Virus in B cells and epithelial cells. *Immunity* 50, 1305.e6–1316.e6. doi: 10.1016/j.immuni.2019.03.010

Chen, J., and Longnecker, R. (2019). Epithelial cell infection by Epstein-Barr virus. *FEMS Microbiol. Rev.* 43, 674–683. doi: 10.1093/femsre/fuz023

Chen, J., Sathiyamoorthy, K., Zhang, X., Schaller, S., Perez White, B. E., Jardetzky, T. S., et al. (2018). Ephrin receptor A2 is a functional entry receptor for Epstein–Barr virus. *Nat. Microbiol.* 3, 172–180. doi: 10.1038/s41564-017-0081-7

Chen, J., Zhang, X., Jardetzky, T. S., and Longnecker, R. (2014). The Epstein-Barr virus (EBV) glycoprotein B cytoplasmic C-terminal tail domain regulates the energy requirement for EBV-induced membrane fusion. *J. Virol.* 88, 11686–11695. doi: 10.1128/JVI.01349-14

Chesnokova, L. S., and Hutt-Fletcher, L. M. (2011). Fusion of Epstein-Barr virus with epithelial cells can be triggered by αvβ5 in addition to αvβ6 and αvβ8, and integrin binding triggers a conformational change in glycoproteins gHgL. *J. Virol.* 85, 13214–13223. doi: 10.1128/JVI.05580-11

Chesnokova, L. S., Valencia, S. M., and Hutt-Fletcher, L. M. (2016). The BDLF3 gene product of Epstein-Barr virus, gp150, mediates non-productive binding to heparan sulfate on epithelial cells and only the binding domain of CD21 is required for infection. *Virology* 494, 23–28. doi: 10.1016/j.virol.2016.04.002

Chia, W. K., Wang, W.-W., Teo, M., Tai, W. M., Lim, W. T., Tan, E. H., et al. (2012). A phase II study evaluating the safety and efficacy of an adenovirus-ΔLMP1-LMP2 transduced dendritic cell vaccine in patients with advanced metastatic nasopharyngeal carcinoma. *Ann. Oncol.* 23, 997–1005. doi: 10.1093/annonc/mdr341

Cohen, J. I. (2015). Epstein–barr virus vaccines. *Clin. Transl. Immunol.* 4:e32. doi: 10.1038/cti.2014.27

Cohen, J. I. (2018). "Vaccine development for Epstein-Barr Virus," in *Human Herpesviruses Advances in Experimental Medicine and Biology*, eds Y. Kawaguchi, Y. Mori, and H. Kimura (Singapore: Springer Singapore), 477–493.

Cohen, J. I., Fauci, A. S., Varmus, H., and Nabel, G. J. (2011). Epstein-Barr virus: an important vaccine target for cancer prevention. *Sci. Transl. Med.* 3:107fs7. doi: 10.1126/scitranslmed.3002878

Connolly, S. A., Jackson, J. O., Jardetzky, T. S., and Longnecker, R. (2011). Fusing structure and function: a structural view of the herpesvirus entry machinery. *Nat. Rev. Microbiol.* 9, 369–381. doi: 10.1038/nrmicro2548

Cui, X., Cao, Z., Chen, Q., Arjunaraja, S., Snow, A. L., and Snapper, C. M. (2016). Rabbits immunized with Epstein-Barr virus gH/gL or gB recombinant proteins elicit higher serum virus neutralizing activity than gp350. *Vaccine* 34, 4050–4055. doi: 10.1016/j.vaccine.2016.06.021

Cui, X., Cao, Z., Ishikawa, Y., Cui, S., Imadome, K.-I., and Snapper, C. M. (2021). Immunization with Epstein–Barr virus core fusion machinery envelope proteins elicit high titers of neutralizing activities and protect humanized mice from lethal dose EBV challenge. *Vaccines* 9:285. doi: 10.3390/vaccines9030285

Cui, X., Cao, Z., Sen, G., Chattopadhyay, G., Fuller, D. H., Fuller, J. T., et al. (2013). A novel tetrameric gp3501–470 as a potential Epstein–Barr virus vaccine. *Vaccine* 31, 3039–3045. doi: 10.1016/j.vaccine.2013.04.071

Dasari, V., Bhatt, K. H., Smith, C., and Khanna, R. (2017). Designing an effective vaccine to prevent Epstein-Barr virus-associated diseases: challenges and opportunities. *Expert Rev. Vaccines* 16, 377–390. doi: 10.1080/14760584.2017.1293529

Dasari, V., Sinha, D., Neller, M. A., Smith, C., and Khanna, R. (2019). Prophylactic and therapeutic strategies for Epstein–Barr virus-associated diseases: emerging strategies for clinical development. *Expert Rev. Vaccines* 18, 457–474. doi: 10.1080/14760584.2019.1605906

Dunmire, S. K., Verghese, P. S., and Balfour, H. H. (2018). Primary Epstein-Barr virus infection. *J. Clin. Virol.* 102, 84–92. doi: 10.1016/j.jcv.2018.03.001

Elliott, S. L., Suhrbier, A., Miles, J. J., Lawrence, G., Pye, S. J., Le, T. T., et al. (2008). Phase I trial of a CD8+ T-Cell peptide epitope-based vaccine for infectious mononucleosis. *J. Virol.* 82, 1448–1457. doi: 10.1128/JVI.01409-07

Epstein, M. A., Achong, B. G., and Barr, Y. M. (1964). Virus particles in cultured lymphoblasts from Burkitt's Lymphoma. *Lancet Lond. Engl.* 1, 702–703. doi: 10.1016/s0140-6736(64)91524-7

Escalante, G. M., Foley, J., Mutsvunguma, L. Z., Rodriguez, E., Mulama, D. H., Muniraju, M., et al. (2020). A pentavalent Epstein-Barr virus-like particle vaccine elicits high titers of neutralizing antibodies against Epstein-Barr virus infection in immunized rabbits. *Vaccines* 8:169. doi: 10.3390/vaccines8020169

Gore, M., and Hutt-Fletcher, L. M. (2009). The BDLF2 protein of Epstein–Barr virus is a type II glycosylated envelope protein whose processing is dependent on coexpression with the BMRF2 protein. *Virology* 383, 162–167. doi: 10.1016/j.virol.2008.10.010

Gu, S. Y., Huang, T. M., Ruan, L., Miao, Y. H., Lu, H., Chu, C. M., et al. (1995). First EBV vaccine trial in humans using recombinant vaccinia virus expressing the major membrane antigen. *Dev. Biol. Stand.* 84, 171–177.

Hartlage, A. S., Liu, T., Patton, J. T., Garman, S. L., Zhang, X., Kurt, H., et al. (2015). The Epstein–Barr Virus lytic protein BZLF1 as a candidate target antigen for vaccine development. *Cancer Immunol. Res.* 3, 787–794. doi: 10.1158/2326-6066.CIR-14-0242

Heeke, D. S., Lin, R., Rao, E., Woo, J. C., McCarthy, M. P., and Marshall, J. D. (2016). Identification of GLA/SE as an effective adjuvant for the induction of robust humoral and cell-mediated immune responses to EBV-gp350 in mice and rabbits. *Vaccine* 34, 2562–2569. doi: 10.1016/j.vaccine.2016.04.012

Heldwein, E. E. (2016). gH/gL supercomplexes at early stages of herpesvirus entry. *Curr. Opin. Virol.* 18, 1–8. doi: 10.1016/j.coviro.2016.01.010

Holmes, D. (2014). The cancer-virus cures. *Nat. Med.* 20, 571–574. doi: 10.1038/nm0614-571

Hui, E. P., Taylor, G. S., Jia, H., Ma, B. B. Y., Chan, S. L., Ho, R., et al. (2013). Phase I trial of recombinant modified vaccinia ankara encoding Epstein-Barr viral tumor antigens in nasopharyngeal carcinoma patients. *Cancer Res.* 73, 1676–1688. doi: 10.1158/0008-5472.CAN-12-2448

Iizasa, H., Kim, H., Kartika, A. V., Kanehiro, Y., and Yoshiyama, H. (2020). Role of viral and host microRNAs in immune regulation of Epstein-Barr virus-associated diseases. *Front. Immunol.* 11:367. doi: 10.3389/fimmu.2020.00367

Johannsen, E., Luftig, M., Chase, M. R., Weicksel, S., Cahir-McFarland, E., Illanes, D., et al. (2004). Proteins of purified Epstein-Barr virus. *Proc. Natl. Acad. Sci. U.S.A.* 101, 16286–16291. doi: 10.1073/pnas.0407320101

Kanekiyo, M., Bu, W., Joyce, M. G., Meng, G., Whittle, J. R. R., Baxa, U., et al. (2015). Rational design of an Epstein-Barr virus vaccine targeting the receptor-binding site. *Cell* 162, 1090–1100. doi: 10.1016/j.cell.2015.07.043

Kang, Y.-F., Zhang, X., Yu, X.-H., Zheng, Q., Liu, Z., Li, J.-P., et al. (2021). Immunization with a Self-assembled nanoparticle vaccine elicits potent neutralizing antibody responses against EBV infection. *Nano Lett.* 21, 2476–2486.

Kenney, S. C., and Mertz, J. E. (2014). Regulation of the latent-lytic switch in Epstein–Barr virus. *Semin. Cancer Biol.* 26, 60–68. doi: 10.1016/j.semcancer.2014.01.002

Khan, G., Fitzmaurice, C., Naghavi, M., and Ahmed, L. A. (2020). Global and regional incidence, mortality and disability-adjusted life-years for Epstein-Barr virus-attributable malignancies, 1990–2017. *BMJ Open* 10:e037505. doi: 10.1136/bmjopen-2020-037505

Kirschner, A. N., Sorem, J., Longnecker, R., and Jardetzky, T. S. (2009). Structure of Epstein-Barr Virus glycoprotein 42 suggests a mechanism for triggering receptor-activated virus entry. *Structure* 17, 223–233. doi: 10.1016/j.str.2008.12.010

Lake, C. M., and Hutt-Fletcher, L. M. (2000). Epstein-Barr virus that lacks glycoprotein gN is impaired in assembly and infection. *J. Virol.* 74, 11162–11172. doi: 10.1128/JVI.74.23.11162-11172.2000

Levin, L. I., Munger, K. L., O'Reilly, E. J., Falk, K. I., and Ascherio, A. (2010). Primary infection with the epstein-barr virus and risk of multiple sclerosis. *Ann. Neurol.* 67, 824–830. doi: 10.1002/ana.21978

Lin, C.-L., Lo, W.-F., Lee, T.-H., Ren, Y., Hwang, S.-L., Cheng, Y.-F., et al. (2002). Immunization with Epstein-Barr Virus (EBV) peptide-pulsed dendritic cells induces functional CD8+ T-cell immunity and may lead to tumor regression in patients with EBV-positive nasopharyngeal carcinoma. *Cancer Res.* 62, 6952–6958.

Lupo, J., Epaulard, O., Morand, P., and Germi, R. (2019). "Le virus d'Epstein-Barr," in *Traité de Virologie Médicale*, 2nd Edn, eds T. Mourez, S. Burrel, D. Boutolleau, and S. Pillet (Paris: Société Française de Microbiologie), 225–237.

Ma, S.-D., Hegde, S., Young, K. H., Sullivan, R., Rajesh, D., Zhou, Y., et al. (2011). A new model of Epstein-Barr Virus infection reveals an important role for early lytic viral protein expression in the development of lymphomas. *J. Virol.* 85, 165–177. doi: 10.1128/JVI.01512-10

Mackett, M., Conway, M. J., Arrand, J. R., Haddad, R. S., and Hutt-Fletcher, L. M. (1990). Characterization and expression of a glycoprotein encoded by the Epstein-Barr virus BamHI I fragment. *J. Virol.* 64, 2545–2552. doi: 10.1128/JVI.64.6.2545-2552.1990

Matsuura, H., Kirschner, A. N., Longnecker, R., and Jardetzky, T. S. (2010). Crystal structure of the Epstein-Barr virus (EBV) glycoprotein H/glycoprotein L (gH/gL) complex. *Proc. Natl. Acad. Sci. U.S.A.* 107, 22641–22646. doi: 10.1073/pnas.1011806108

Miller, N., and Hutt-Fletcher, L. M. (1992). Epstein-Barr virus enters B cells and epithelial cells by different routes. *J. Virol.* 66, 3409–3414. doi: 10.1128/JVI.66.6.3409-3414.1992

Möhl, B. S., Chen, J., and Longnecker, R. (2019). Gammaherpesvirus entry and fusion: a tale how two human pathogenic viruses enter their host cells. *Adv. Virus Res.* 104, 313–343. doi: 10.1016/bs.aivir.2019.05.006

Möhl, B. S., Chen, J., Park, S. J., Jardetzky, T. S., and Longnecker, R. (2017). Epstein-Barr Virus fusion with epithelial cells triggered by gB is restricted by a gL glycosylation site. *J. Virol.* 91, e1255–e1317. doi: 10.1128/JVI.01255-17

Mohl, B. S., Chen, J., Sathiyamoorthy, K., Jardetzky, T. S., and Longnecker, R. (2016). Structural and mechanistic insights into the tropism of Epstein-Barr Virus. *Mol. Cells* 39, 286–291. doi: 10.14348/MOLCELLS.2016.0066

Moutschen, M., Léonard, P., Sokal, E. M., Smets, F., Haumont, M., Mazzu, P., et al. (2007). Phase I/II studies to evaluate safety and immunogenicity of a recombinant gp350 Epstein–Barr virus vaccine in healthy adults. *Vaccine* 25, 4697–4705. doi: 10.1016/j.vaccine.2007.04.008

Mui, U. N., Haley, C. T., Vangipuram, R., and Tyring, S. K. (2019). Human oncoviruses: mucocutaneous manifestations, pathogenesis, therapeutics, and prevention. *J. Am. Acad. Dermatol.* 81, 23–41. doi: 10.1016/j.jaad.2018.10.072

Münz, C. (2019). Latency and lytic replication in Epstein–Barr virus-associated oncogenesis. *Nat. Rev. Microbiol.* 17, 691–700. doi: 10.1038/s41579-019-0249-7

Münz, C. (2020). Redirecting T cells against Epstein–Barr Virus Infection and associated oncogenesis. *Cells* 9:1400. doi: 10.3390/cells9061400

Niederman, J. C., and Evans, A. S. (1997). "Epstein-Barr virus," in *Viral Infections of Humans: Epidemiology and Control*, 4th Edn, eds A. S. Evans and R. A. Kaslow (New York, NY: Plenum Medical Book Co.), 253–283.

Niedobitek, G. (1999). The Epstein-Barr virus: a group 1 carcinogen? *Virchows Arch.* 435, 79–86. doi: 10.1007/s004280050402

Pavlova, S., Feederle, R., Gartner, K., Fuchs, W., Granzow, H., and Delecluse, H.-J. (2013). An Epstein-Barr virus mutant produces immunogenic defective particles devoid of viral DNA. *J. Virol.* 87, 2011–2022. doi: 10.1128/JVI.02533-12

Pei, Y., Wong, J. H. Y., and Robertson, E. S. (2020). Targeted therapies for Epstein-Barr Virus-associated lymphomas. *Cancers* 12:2565. doi: 10.3390/cancers12092565

Perez, E. M., Foley, J., Tison, T., Silva, R., and Ogembo, J. G. (2017). Novel Epstein-Barr virus-like particles incorporating gH/gL-EBNA1 or gB-LMP2 induce high neutralizing antibody titers and EBV-specific T-cell responses in immunized mice. *Oncotarget* 8, 19255–19273. doi: 10.18632/oncotarget.13770

Rees, L., Tizard, E. J., Morgan, A. J., Cubitt, W. D., Finerty, S., Oyewole-Eletu, T. A., et al. (2009). A Phase I trial of Epstein-Barr Virus Gp350 vaccine for children with chronic kidney disease awaiting transplantation. *Transplantation* 88, 1025–1029. doi: 10.1097/TP.0b013e3181b9d918

Rickinson, A. B., and Kieff, E. (2013). "Epstein-Barr Virus," in *Fields Virology*, 6th Edn, eds D. M. Knipe and P. M. Howley (New York, NY: Lippincott, Williams and Wilkins), 655–700. Editors-in-chief, D. M. Knipe, P. M. Howley, associate editors, J. I. Cohen.

Rühl, J., Citterio, C., Engelmann, C., Haigh, T., Dzionek, A., Dreyer, J., et al. (2019). Heterologous prime-boost vaccination protects against EBV antigen–expressing lymphomas. *J. Clin. Invest.* 129, 2071–2087. doi: 10.1172/JCI125364

Rühl, J., Leung, C. S., and Münz, C. (2020). Vaccination against the Epstein–Barr virus. *Cell. Mol. Life Sci.* 77, 4315–4324. doi: 10.1007/s00018-020-03538-3

Sathiyamoorthy, K., Hu, Y. X., Möhl, B. S., Chen, J., Longnecker, R., and Jardetzky, T. S. (2016). Structural basis for Epstein–Barr virus host cell tropism mediated by gp42 and gHgL entry glycoproteins. *Nat. Commun.* 7:13557. doi: 10.1038/ncomms13557

Sathiyamoorthy, K., Jiang, J., Hu, Y. X., Rowe, C. L., Möhl, B. S., Chen, J., et al. (2014). Assembly and architecture of the EBV B cell entry triggering complex. *PLoS Pathog.* 10:e1004309. doi: 10.1371/journal.ppat.1004309

Sathiyamoorthy, K., Jiang, J., Möhl, B. S., Chen, J., Zhou, Z. H., Longnecker, R., et al. (2017). Inhibition of EBV-mediated membrane fusion by anti-gHgL antibodies. *Proc. Natl. Acad. Sci. U.S.A.* 114, E8703–E8710. doi: 10.1073/pnas.1704661114

Shannon-Lowe, C., Rickinson, A. B., and Bell, A. I. (2017). Epstein–Barr virus-associated lymphomas. *Philos. Trans. R. Soc. B Biol. Sci.* 372:20160271. doi: 10.1098/rstb.2016.0271

Si, Y., Deng, Z., Lan, G., Du, H., Wang, Y., Si, J., et al. (2016). The safety and immunological effects of rAd5-EBV-LMP2 vaccine in nasopharyngeal carcinoma patients: a Phase I clinical trial and two-year follow-up. *Chem. Pharm. Bull. (Tokyo)* 64, 1118–1123. doi: 10.1248/cpb.c16-00114

Snijder, J., Ortego, M. S., Weidle, C., Stuart, A. B., Gray, M. D., McElrath, M. J., et al. (2018). An antibody targeting the fusion machinery neutralizes dual-tropic infection and defines a site of vulnerability on Epstein-Barr Virus. *Immunity* 48, 799.e9–811.e9. doi: 10.1016/j.immuni.2018.03.026

Sokal, E. M., Hoppenbrouwers, K., Vandermeulen, C., Moutschen, M., Léonard, P., Moreels, A., et al. (2007). Recombinant gp350 vaccine for infectious mononucleosis: a phase 2, randomized, double-blind, placebo-controlled trial to evaluate the safety, immunogenicity, and efficacy of an Epstein-Barr virus

vaccine in healthy young adults. *J. Infect. Dis.* 196, 1749–1753. doi: 10.1086/523813

Sorgato, C. C., Lins-e-Silva, M., Leão, J. C., Vasconcelos, L. R., Romão, T. P., Duarte, A. L., et al. (2020). EBV and CMV viral load in rheumatoid arthritis and their role in associated Sjögren's syndrome. *J. Oral Pathol. Med.* 49, 693–700. doi: 10.1111/jop.13036

Straathof, K. C. M., Bollard, C. M., Popat, U., Huls, M. H., Lopez, T., Morriss, M. C., et al. (2005). Treatment of nasopharyngeal carcinoma with Epstein-Barr virus–specific T lymphocytes. *Blood* 105, 1898–1904. doi: 10.1182/blood-2004-07-2975

Tang, Y., Lu, S., Gan, X., Liu, F., Zhang, Y., Luo, C., et al. (2016). Expression of LMP and EBNA genes in Epstein-Barr virus-associated lymphomas in Hu-PBL/SCID mice. *Oncol. Rep.* 35, 905–911. doi: 10.3892/or.2015.4401

Tangye, S. G., Palendira, U., and Edwards, E. S. J. (2017). Human immunity against EBV—lessons from the clinic. *J. Exp. Med.* 214, 269–283. doi: 10.1084/jem.20161846

Taylor, G. S., Haigh, T. A., Gudgeon, N. H., Phelps, R. J., Lee, S. P., Steven, N. M., et al. (2004). Dual Stimulation of Epstein-Barr Virus (EBV)-specific CD4+- and CD8+-T-cell responses by a chimeric antigen construct: potential therapeutic vaccine for EBV-positive nasopharyngeal carcinoma. *J. Virol.* 78, 768–778. doi: 10.1128/JVI.78.2.768-778.2004

Taylor, G. S., Jia, H., Harrington, K., Lee, L. W., Turner, J., Ladell, K., et al. (2014). A recombinant modified vaccinia ankara vaccine encoding Epstein-Barr Virus (EBV) target antigens: a Phase I trial in UK patients with EBV-positive cancer. *Clin. Cancer Res.* 20, 5009–5022. doi: 10.1158/1078-0432.CCR-14-1122-T

van Zyl, D. G., Mautner, J., and Delecluse, H.-J. (2019). Progress in EBV Vaccines. *Front. Oncol.* 9:104. doi: 10.3389/fonc.2019.00104

van Zyl, D. G., Tsai, M.-H., Shumilov, A., Schneidt, V., Poirey, R., Schlehe, B., et al. (2018). Immunogenic particles with a broad antigenic spectrum stimulate cytolytic T cells and offer increased protection against EBV infection ex vivo and in mice. *PLoS Pathog.* 14:e1007464. doi: 10.1371/journal.ppat.1007464

Walling, D. M., Brown, A. L., Etienne, W., Keitel, W. A., and Ling, P. D. (2003). Multiple Epstein-Barr virus infections in healthy individuals. *J. Virol.* 77, 6546–6550. doi: 10.1128/JVI.77.11.6546-6550.2003

Wang, H.-B., Zhang, H., Zhang, J.-P., Li, Y., Zhao, B., Feng, G.-K., et al. (2015). Neuropilin 1 is an entry factor that promotes EBV infection of nasopharyngeal epithelial cells. *Nat. Commun.* 6:6240. doi: 10.1038/ncomms7240

Zhang, X., Zhao, B., Ding, M., Song, S., Kang, Y., Yu, Y., et al. (2020). A novel vaccine candidate based on chimeric virus-like particle displaying multiple conserved epitope peptides induced neutralizing antibodies against EBV infection. *Theranostics* 10, 5704–5718. doi: 10.7150/thno.42494

Zhao, B., Zhang, X., Krummenacher, C., Song, S., Gao, L., Zhang, H., et al. (2018). Immunization with Fc-based recombinant Epstein–Barr virus gp350 elicits potent neutralizing humoral immune response in a BALB/c mice model. *Front. Immunol.* 9:932. doi: 10.3389/fimmu.2018.00932

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Microbiology

ORIGINAL RESEARCH
published: 09 July 2021
doi: 10.3389/fmicb.2021.677847

Generation of Infectious Mimivirus Virions Through Inoculation of Viral DNA Within *Acanthamoeba castellanii* Shows Involvement of Five Proteins, Essentially Uncharacterized

*Dehia Sahmi-Bounsiar[1,2], Jean-Pierre Baudoin[1,2], Sihem Hannat[1,2], Philippe Decloquement[1,2], Eric Chabrieres[1,2], Sarah Aherfi[1,2] and Bernard La Scola[1,2]**

[1] IHU Méditerranée Infection, Marseille, France, [2] Aix-Marseille Université, Institut de Recherche pour le Développement (IRD), Assistance Publique- Hôpitaux de Marseille (AP-HM), MEPHI, Marseille, France

OPEN ACCESS

Edited by:
Masaharu Takemura, Tokyo University of Science, Japan

Reviewed by:
Kenta Okamoto, Uppsala University, Sweden
Patrick Forterre, Institut Pasteur, France

***Correspondence:**
Bernard La Scola
bernard.la-scola@univ-amu.fr

Specialty section:
This article was submitted to Virology, a section of the journal Frontiers in Microbiology

Received: *08 March 2021*
Accepted: *08 June 2021*
Published: *09 July 2021*

Citation:
Sahmi-Bounsiar D, Baudoin J-P, Hannat S, Decloquement P, Chabrieres E, Aherfi S and La Scola B (2021) Generation of Infectious Mimivirus Virions Through Inoculation of Viral DNA Within Acanthamoeba castellanii Shows Involvement of Five Proteins, Essentially Uncharacterized. Front. Microbiol. 12:677847. doi: 10.3389/fmicb.2021.677847

One of the most curious findings associated with the discovery of *Acanthamoeba polyphaga* mimivirus (APMV) was the presence of many proteins and RNAs within the virion. Although some hypotheses on their role in Acanthamoeba infection have been put forward, none have been validated. In this study, we directly transfected mimivirus DNA with or without additional proteinase K treatment to extracted DNA into *Acanthamoeba castellanii.* In this way, it was possible to generate infectious APMV virions, but only without extra proteinase K treatment of extracted DNA. The virus genomes before and after transfection were identical. We searched for the remaining DNA-associated proteins that were digested by proteinase K and could visualize at least five putative proteins. Matrix-assisted laser desorption/ionization time-of-flight and liquid chromatography–mass spectrometry comparison with protein databases allowed the identification of four hypothetical proteins—L442, L724, L829, and R387—and putative GMC-type oxidoreductase R135. We believe that L442 plays a major role in this protein–DNA interaction. In the future, expression in vectors and then diffraction of X-rays by protein crystals could help reveal the exact structure of this protein and its precise role.

Keywords: L442, single-cell transfection, microinjection, *Acanthamoeba castellanii*, ApMV, L724, L829, R387

INTRODUCTION

Acanthamoeba castellanii (Byers, 1979) is a small, complex, and free-living amoeba that can also live as a parasite within a host tissue. In some cases, it is associated with human diseases (Król-turmińska and Olender, 2017). *Acanthamoeba* has been a useful model in various biological studies (Weisman and Korn, 1967; Read and Kabana, 1980; Fouque et al., 2012; La Scola, 2014; Sahmi-bounsiar et al., 2019; Hasni et al., 2019), notably for its membrane capacity to engulf molecules and/or microorganisms whether naturally by phagocytosis of bacteria and viruses (Weisman and Korn, 1967; La Scola et al., 2001; Araú et al., 2016; Raoult and Boyer, 2010) or artificially *via* chemical

(Peng et al., 2005; Mougari et al., 2019) or mechanical (Sinard and Pollard, 1989) transfection processes for antibodies, genomic DNA, plasmids, or fluorochrome delivery. Transfection methods have been used for decades (Eisenstark, 1965; Kim and Eberwine, 2010), and among physical transfections, microinjection is a technique that enables the integration of cells or large molecules on a microscopic scale in cells, whether adherent or in suspension culture (Wang et al., 2008; Dean and Gasiorowski, 2010). It has revolutionized the medical field by making *in vitro* fertilization possible (Lopata et al., 1980). In biology, microinjection was one of the first transfection tools used for the study of several cellular processes (Zhang and Yu, 2008). The microinjection of *A. castellanii* has only been carried out once, in 1989, with the aim of studying the motility of this amoeba (Sinard and Pollard, 1989). Due to its complexity, the need for special equipment (Dean et al.) and skilled experimenters, the use of microinjection regressed with the advent of chemical transfection.

With the improvement of the coculture process on amoeba (Yaacoub et al., 2016), our laboratory has greatly contributed to the discovery and improvement of the isolation of giant viruses of amoeba (Rolland et al., 2019) since the discovery of *Acanthamoeba polyphaga* mimivirus (APMV) in 2003 (La Scola et al., 2003). One of the most curious findings associated with the discovery of this virus was the presence of many proteins and RNA within the virion (Suzan-Monti et al., 2007). The involvement of these RNA and proteins has been suggested to be associated with the early stages of infection but has never been fully investigated. The aim of this study was to explore this hypothesis by directly transfecting APMV DNA into *A. castellanii.* Using microinjection, we were able to transfect *A. castellanii* amoeba with mimivirus extracted DNA and generate infectious APMV virions. We revealed the need for DNA-mediated APMV generation of at least four uncharacterized proteins—L442, L724, L829, and R387—and putative GMC-type oxidoreductase R135.

MATERIALS AND METHODS

Cell Preparation

We used *A. castellanii* (ATCC 30010) as a cellular support in peptone–yeast extract–glucose (PYG) medium at a concentration of 5×10^5 cells/ml cultured at 28°C in 75-cm^2 cell culture flasks, as previously described (Yaacoub et al., 2016). After 48 h of incubation, the flask was gently tapped to detach adherent cells, which were centrifuged for 10 min at 500 × *g* to remove all amoebae debris. The cell pellet was then resuspended and washed twice in starvation medium (Yaacoub et al., 2016). A suspension containing 2 ml of amoebae at 10^3 cells/ml was then plated into a cell imaging dish (Ibidi glass-bottomed 35-mm petri dish; Germany) with low confluence, allowing for good observation and manipulation control.

DNA Extraction and Proteinase K Treatment

For mimivirus production, 10 150-cm^2 flasks containing 10 ml of *A. castellanii* at 5×10^5 cells/ml in 30 ml of PYG were inoculated with 5 ml of *A. polyphaga* mimivirus at a multiplicity of infection of 10. The cocultures were incubated at 30°C and checked daily by inverted optical microscopy to observe cytopathic effects. After the complete lysis of the amoeba cells, the virus supernatant was collected from the cultures and then filtered through 0.8-μm-pore filters to eliminate debris. The supernatant was centrifuged at 14,000 × *g* for 45 min. The supernatant was removed by aspiration, and the pellet was then resuspended in 1 ml of phosphate-buffered saline. The virus was then purified by ultracentrifugation at 14,000 × *g* for 45 min across a 25% sucrose layer, and the viral pellet was resuspended with 1 ml of phosphate-buffered saline and stored at −80°C. Viral DNA was then extracted from 200 μl of the purified virus (10^8 particles/ml) using the EZ1 advanced XL and using EZ1 DNA Tissue Kit (Qiagen, Hilden, Germany) according to the manufacturer's instructions. Extracted DNA was doubly filtrated through a 0.22-μm-pore filter and quantified with NanoDrop™ 2000 at approximately 150 ng/ μl. The extracted APMV DNA concentration was diluted to 10 ng/ml for microinjection. To remove the remaining proteins, proteinase K (Thermo Fisher Scientific, United States) treatment was performed by adding to 200 μl of APMV DNA, 200 μl of Tampon G2, and 10 μl of proteinase K. The digestion process takes place at 56°C for 2 h. This second digestion was thereafter referred to as proteinase K extracted DNA pre-treatment.

Microinjection Components and Procedure

The workstation (**Figure 1**) essentially comprises an injectMan NI2 micromanipulator (Eppendorf Equipment, France), which allows the micropipette to be positioned, as well as a femtoJet 4i microinjector (Eppendorf, France), micropipettes Femtotips II (0.5 μm inner diameter and 0.7 μm outer diameter) (Eppendorf, France), an eclipse TE 2000S inverted microscope (Nikon, France), and a DFC 425C camera (Leica, Germany). A computer module was used to observe manipulations and take pictures (Nikon, France). For microinjection, the femtoJet 4i system was used at an injection pressure (Pi) of 75 hectopascals (hPa), a compensation (holding) pressure (Pc) of 10 hPa, and 0.2 s for the time injection (Ti). The microinjection solution was composed of 1 μl of a red fluorescent dye (Dextran Rhodamine B, 70,000 molecular weight, neutral, Invitrogen) and 9 μl of APMV DNA extract at a concentration of 10 ng/μl before injection amoebas were placed into the cell imaging dish (Eppendorf) with 2 ml of starvation medium. Using a microloader, the microinjection needle was filled with 2 μl of the injection solution and then mounted onto an Eppendorf micropipette holder attached to an eclipse TE 2000S inverted microscope (Nikon, France) with an epifluorescence system. The cells were injected by maintaining a constant low flow rate out of the needle tip. The needle was inserted into the amoeba at a shallow angle of −45°, kept in the cell enough to inject about 5–10% of the cell volume, and then removed. Usually, the microinjected volume is on the order of a femtoliter or picoliter. The microinjection volume is determined by the injection parameters (Pi and Ti), the type of femtotips (its opening and shape), and the viscosity of the microinjected

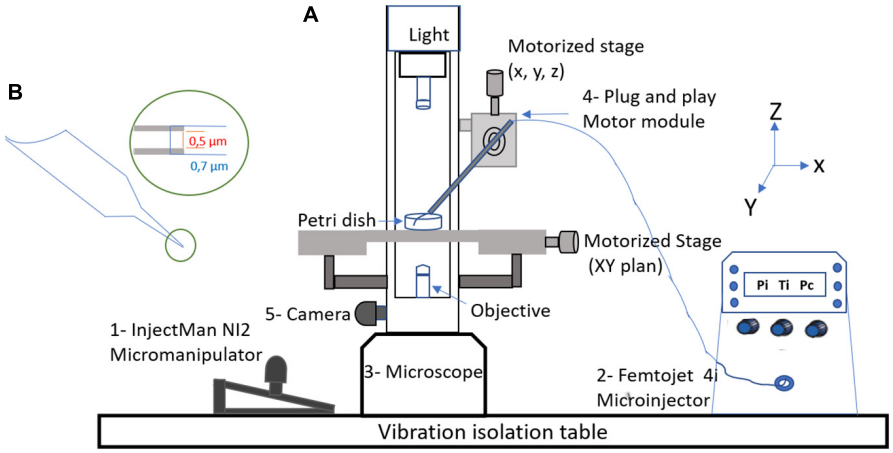


FIGURE 1 | Micromanipulation equipment. **(A)** Components of workstation: 1—InjectMan NI2 micromanipulator which allows for the positioning of the micropipette (Eppendorf Equipment, France), 2—FemtoJet 4i microinjector (Eppendorf, France), 3—Eclipse TE 2000S Inverted microscope (Nikon, France), 4—plug and play motor modules, and 5—DFC 425C camera (Leica, Germany). **(B)** Femtotips with an inner diameter of 0.5 µm.

solution. Only indirect methods can determine the approximate volume as performed in this study (Keith et al., 1983).

Retention of the fluorescent label was indicative of successful cell microinjection. During each session of microinjection, a negative control consisting of amoeba that were not microinjected was performed by adding 9 µl of APMV DNA extract and 1 µl of red fluorescent dye to fresh amoeba in 2 ml of starvation medium. Afterward, the successfully microinjected amoebas were monitored microscopically for the assessment of the presence of APMV virions released into the medium. The APMV virions could be found after a period of between 1 and 3 weeks, during which the culture medium was changed regularly. Once the presence of virions was noted, cells were scraped from the dish and subcultured along with the resulting culture supernatant in a new petri dish containing a monolayer of fresh amoeba.

Flow Cytometry Detection

Flow cytometry based on side scatter and DNA content was used. After cytopathic effect and lysis detection, a supernatant was centrifuged at 700 × *g* for 10 min to discard large debris. The supernatant was stained using SYBR green dye (SYBR green I nucleic acid gel stain; Molecular Probes, Life Technologies, Inc., Carlsbad, CA, United States) at a dilution of 1:100 and heated to 80°C for 3 min. Data were collected on a BD LSR Fortessa (BD Biosciences) cytometer and compared with those for previously known gated viruses by using FlowJo software.

Sample Preparation and Image Acquisition for Scanning Electron Microscopy

The samples were centrifuged at 14,000 × *g* for 10 min, and the supernatant was suspended in 2.5% of glutaraldehyde fixative solution. We then directly placed the sample onto microscopy slides for observation. We used the Hitachi TM4000 scanning electron microscope (SEM) (Hitachi, Japan) for image acquisition.

Comparative Genomic Analysis

Genomic DNA mimivirus (pre-microinjection and post-microinjection, respectively) was quantified using a Qubit assay with the high sensitivity kit (Life Technologies, Carlsbad, CA, United States) to 0.2 ng/µl. The genomic DNA was then sequenced on MiSeq Technology (Illumina Inc., San Diego, CA, United States) with the paired end strategy and was barcoded in order to be mixed, respectively, with 28 other genomic projects prepared with the Nextera XT DNA sample prep kit (Illumina). Dilution was performed to reach 1 ng of each genome as input to prepare the paired end library. The "tagmentation" step fragmented and tagged the DNA. The limited-cycle PCR amplification (12 cycles) then completed the tag adapters and introduced dual-index barcodes. After purification on AMPure XP beads (Beckman Coulter Inc., Fullerton, CA, United States), the libraries were then normalized on specific beads according to the Nextera XT protocol (Illumina). Normalized libraries were pooled into a single library for sequencing on the MiSeq. The pooled single-strand library was loaded onto the reagent cartridge and then onto the instrument along with the flow cell. Automated cluster generation and paired end sequencing with dual index reads were performed in a single 39-h run in 2 × 250 bp. Total information of 12.47 Gb was obtained from a 642-K/mm^2 cluster density, with a cluster passing quality control filters of 95.17%. Within this run, the index representation for mimivirus pre-microinjection was determined to be 3.05%, and it was also 3.66% for mimivirus post-microinjection. The 11,864,976 paired end reads were filtered according to the read quality. To compare the genetic contents between both isolates (before and after microinjection), all annotations were performed using the Prokka annotation pipeline (Seemann, 2014) and then used in the Roary pan-genome pipeline using default parameters (Page et al., 2015). The resulting core genome

alignments were used to study SNPs using SNP-Sites (Page et al., 2016). These results were double-checked by an in-house script for detecting SNPs.

Identification of DNA-Associated Proteins

We used both silver and Coomassie blue staining originally developed to detect proteins separated by SDS-PAGE (Switzer, 1979; Merril et al., 1981; Heukeshoven and Dernick, 1985).

Five protein bands were excised manually from Coomassie Blue staining gels. After several successive washes with acetonitrile and water, in-gel digestion with proteomics-grade trypsin (Agilent Technologies) was done overnight at room temperature. The peptides obtained from protein digestion were extracted with acetonitrile.

Peptides were identified as a first step using MALDI-TOF-MS (matrix-assisted laser desorption/ionization time-of-flight mass spectrometry) on a Bruker Autoflex Speed spectrometer (Bruker Daltonics) and as a second step using a nanoAcquity UPLC system connected to a Synapt G2Si Q-TOF spectrometer (Waters).

For the MALDI-TOF analyses, 1 μl peptide mixture was cocrystallized onto the Anchorchip MALDI-TOF target plate with an equal amount of matrix solution (0.3 mg/ml of α-cyano-4-hydroxycinnamic acid in acetone/ethanol, 1:2 v/v, acidified with TFA, 0.1% final). The mass spectrometer was calibrated externally using bovine serum albumin tryptic peptides. The peptide mass fingerprints were used to identify the proteins.

For the liquid chromatography–mass spectrometry (LC-MS) analyses, peptides were pooled and were eluted onto a trapping column (nanoAcquity UPLC 2G-V/M Trap 5μm Symmetry C18 180 μm × 20 mm, Waters) for concentration and desalting at 10 μl/min of 99.9% water, 0.1% formic acid, 0.1% acetonitrile, and 0.1% formic acid. The peptides were eluted on a C18 100 μm × 100 mm column (nanoAcquity UPLC 1.7 μm BEH C18, Waters) and separated using a 100-min gradient (300 nl/min, 5–40% acetonitrile, and 0.1% formic acid). Data-dependent MS/MS monitoring was performed in positive mode. GFP lock mass correction was applied to spectra. Raw MS data was processed using PEAKS Studio 6.0 software. Swissprot online protein sequences were used for protein identification. Proteins presenting one or more peptides were considered as identified.

A tertiary structure prediction was used for all the uncharacterized proteins—L442, L724, L829, and R387—using Phyre2 tool (Kelley et al., 2015).

RESULTS

Production of APMV Virions After Microinjection of APMV DNA in Amoeba

Our microinjection methodology was successful in a quarter of the 200 sessions performed. Successful experiments made it possible to achieve between one and eight microinjected amoebae, as checked by fluorescent-dextran loading of the cells (**Figures 2**, **4**). Of these 50 successful experiments, six led to amoeba monolayer lysis associated with the production of mimivirus virions. Despite the shock reaction with cell rounding and detachment from the substrate that was initially observed optically (**Figure 2**), the successfully microinjected cells recognized by the fluorescent cytoplasm recovered their normal morphology within 1–2 h after microinjection (**Figure 3**). A confirmation of cell morphology, motility, and viability was performed 24 h post-microinjection (**Supplementary Video 1**). The viability, as checked by observing cell morphology and motility, was comparable to non-microinjected cells. None of the negative controls led to amoeba lysis. Of the six microinjections leading to the production of viral particles, a cytopathic effect consisting of slight lysis or rounding of a fraction of the amoeba population was observed, starting from 5 to 7 days post-microinjection. A subculture of the supernatant and cells of these dishes on fresh amoeba was then carried out and followed by daily optical observation. Amoeba lysis at the second passage was observed after 2–4 days, with the production of viral particles as detected first with optical microscopy and then confirmed by SEM and flow cytometry. SEM (Hitachi TM4000) showed viral particles presenting the same morphological characteristics as APMV and with mean maximal diameters between 460 and 500 nm (**Figure 4B**). Flow cytometry dot plot also confirmed the production of mimivirus, showing in SSC (side scatter) versus FITC (SYBR green DNA contents) a single viral population corresponding to the size/structure profile of mimivirus virions and quantified at 10^8 particles/ml (**Figure 4A**). Genome sequencing was performed on purified mimivirus solution and

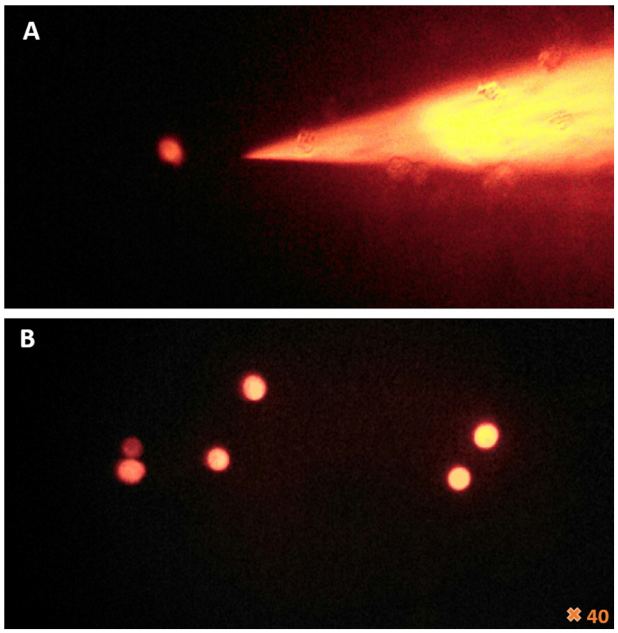


FIGURE 2 | Fluorescence microscopy of the femtotips and the cell's reaction to microinjection. **(A)** Illustration of femtotips removed from microinjected amoeba. **(B)** Observation of amoeba stained in red by rhodamine-dextran and the effect of the microinjection which induces the loss of amoeba adherence and their rounding.

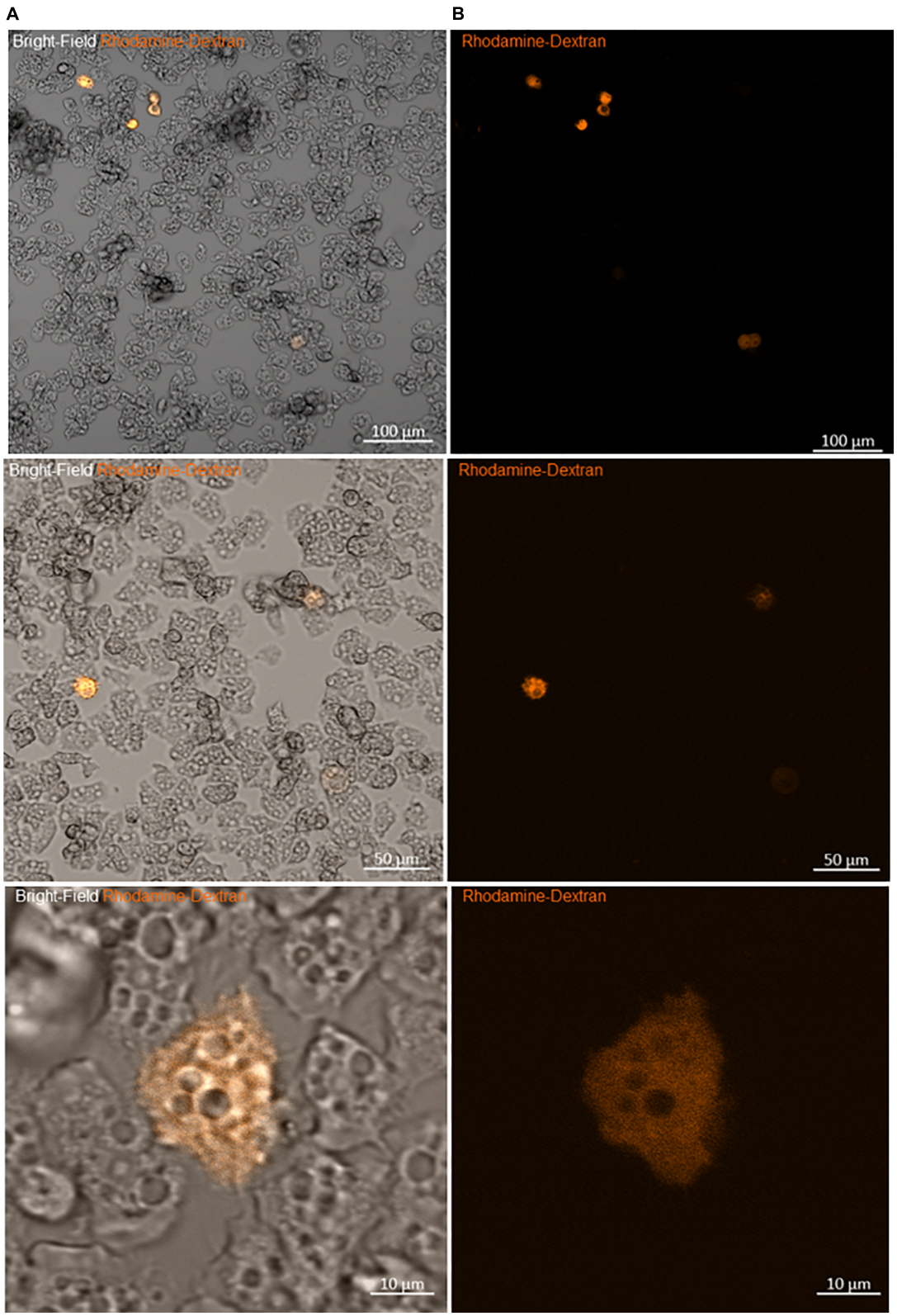


FIGURE 3 | Viability of amoeba microinjected with rhodamine-dextran dye. **(A)** Bright-field images showing the maintenance of the trophozoite state of different amoebae 2 h after microinjection. **(B)** Fluorescence microscopy images of the same frame show the homogeneity of the red dye into the cytoplasm, without any captation by other structures, sign of the successful entry of DNA and the success of the microinjection.

on viral particles produced by amoeba after microinjection. Sequence analysis confirmed that both analyzed genomes were that of *Acanthamoeba polyphaga* mimivirus (GenBank access number AY653733) and revealed 100% similarity between the two viral isolates (**Supplementary Figure 1**).

APMV Virion Production by Microinjected Amoeba Is Blocked by Proteinase K Treatment of APMV DNA and Involves Certain Proteins

To check if some proteins associated with APMV DNA are necessary for viral production, a step of pre-treatment with proteinase K was added to the DNA after its extraction in order to eliminate any residual protein. In parallel, 50 microinjection sessions with pre-treatment with proteinase K and 50 sessions without pre-treatment were carried out. Out of 50 sessions without proteinase K treatment, we obtained 12 successful experiments, each one making possible to achieve between one and five microinjected amoebae. Two of these led to viral production. In contrast, no microinjection with proteinase K pre-treated APMV DNA led to amoeba monolayer lysis associated with APMV particle production.

To better understand the nature of the proteinase K digested material from APMV DNA extract, protein analysis was carried out using SDS-PAGE. This analysis revealed five constant putative protein bands (**Figures 5A,C**). In-gel digestion and MALDI-TOF-MS showed the presence in a single band (band no. 3) of the cleaved sequence of an uncharacterized protein, L442 with a size between 43 and 55,223 kDa (**Figure 5C**). Band

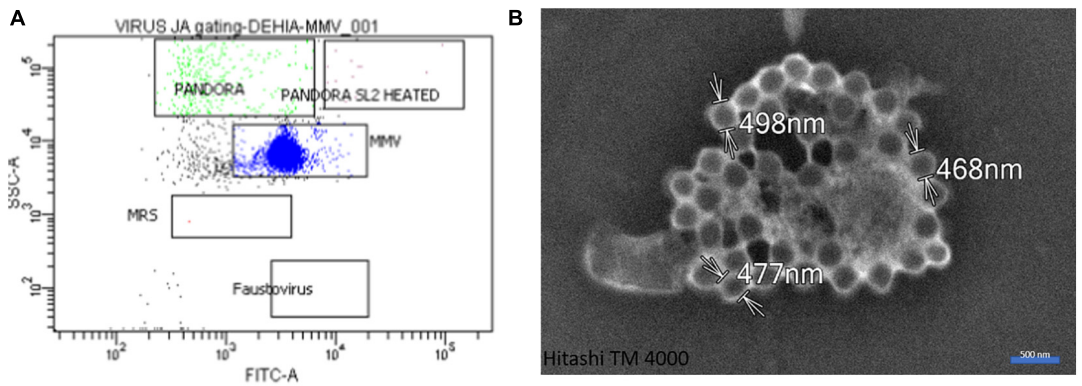


FIGURE 4 | Gating and microscopic control. **(A)** Flow cytometry dot plot showing the single viral profile in SSC (side scatter) versus FITC (SYBR green DNA contents) of mimivirus. **(B)** Scanning electron microscopy with Hitachi TM4000 of a culture supernatant showing mimivirus particles.

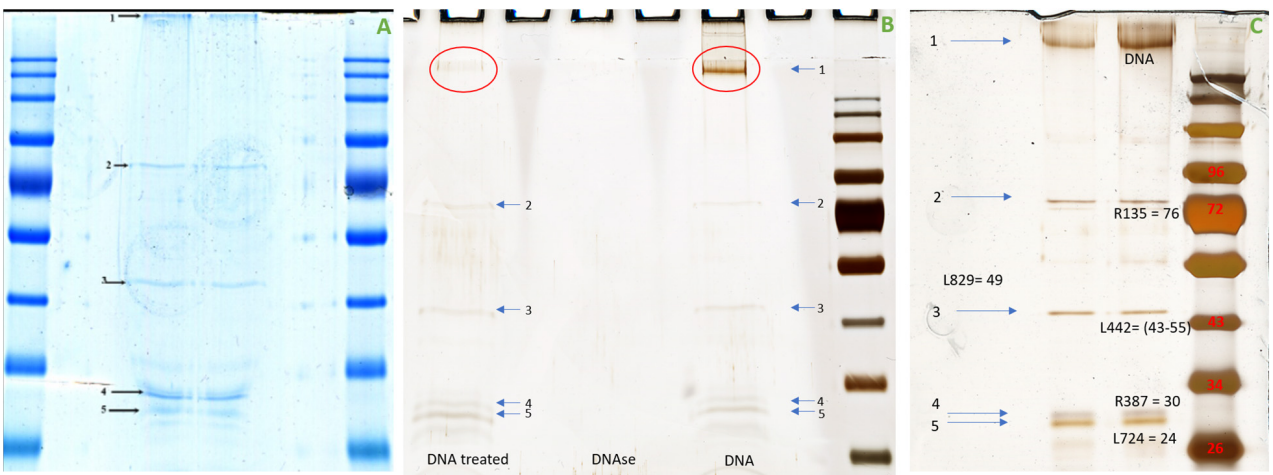


FIGURE 5 | Migration of the APMV DNA extract into polyacrylamide gel electrophoresis. **(A)** Coomassie blue staining (InstantBlue Protein Stain) of polyacrylamide gel electrophoresis showed five different bands possibly corresponding to proteins. **(B)** Silver staining of polyacrylamide gel electrophoresis showed the disappearance of band no. 1 when the APMV DNA is treated with DNase and the persistence of the band nos. 2, 3, 4, and 5. **(C)** Silver staining of polyacrylamide gel electrophoresis showed five bands corresponding, respectively, to DNA, R135 putative GMC-type oxidoreductase (molecular weight = 76,947 Da), cleaved sequence of the hypothetical protein L442 (molecular weight ≃ 48,000 Da), hypothetical protein R387 (molecular weight = 30,067 Da), and hypothetical protein L724 (molecular weight = 24,033 Da). No visible band corresponds to hypothetical protein L829 (molecular weight = 49,229 Da). The protein marker used was Color Prestained Protein Standard, Broad Range (10–250 kDa).

no. 1 corresponds to DNA, as indicated by its removal after DNase treatment, while the other bands persist (**Figure 5B**). The intensities of band nos. 2, 3, 4, and 5 were too low to be analyzed with MALDI-TOF. This result was confirmed by those obtained by LC-MS which give as first hit one L442 (139,334 Da) with 11% coverage for 12 identified peptides. Putative GMC-type oxidoreductase R135 (76,947 Da) was also found with 16% coverage for 10 identified peptides. Other mimivirus proteins could also be identified with one peptide for uncharacterized protein R387 (30,067 Da) and two for uncharacterized proteins L724 (24,033 Da) and L829 (49,226 Da) (**Supplementary Figure 2**).

Homologs of uncharacterized protein L422 were present in lineage A, B, and C of mimiviruses and in tupanviruses. A more distant homolog that we detected by delta Blast is also present in an archaeon, in agreement with an analysis of the relationship between Nucleo-Cytoplasmic Large DNA Viruses (NCLDVs) and cell domains concluding on the emergence of NCLDVs before the last eukaryotic common ancestor (Guglielmini et al., 2019). A phylogenetic analysis of L422 was performed either without outgroup (**Supplementary Figure 5**) or using the archaeal as an outgroup (**Figure 6**). The results suggest that L422 was already present in the common ancestor of Mimiviridae and Tupanvirus.

The proteins L829 and R387 also have homologs in the three lineages of mimiviruses and in tupanviruses, and phylogenetic analysis suggests that they were already present in the common ancestor of these viruses (**Supplementary Figures 5–7**).

The putative GMC-type oxidoreductase R135 and the uncharacterized protein L724 have only homologs in the three lineages of mimivirus, and the phylogenetic analysis suggests that it was already present in the common ancestor of the A, B, and C lineages (**Supplementary Figures 3**, **4**).

Using phyre2 tool, a tertiary structure prediction applied first to the entire sequence of L442 then to the cleaved sequence, showed both similarity with human ATP-dependant DNA helicase q5 (first hit; Confidence: 47.2% Identity 32%) and (first hit; Confidence: 48% Identity 32%), respectively. These scores were too low to predict any putative function, but both organisms are also phylogenetically distant. This result suggests that this protein sequence may be involved in DNA metabolism (**Supplementary Figure 8**).

Using the same tool, a tertiary structure applied to the sequence of L724, L829, and R387 showed a similarity related, respectively, to the field of transcription regulation (first hit; confidence: 65.2%, identity: 42%), hydrolase (first hit; confidence: 91.2%, identity 32%), and cysteine zipper (first hit; confidence: 94.9%, identity: 20%). The scores stay low to predict any putative function (**Supplementary Figure 8**).

DISCUSSION

Our microinjection methodology for the inoculation of APMV DNA extract into *A. castellanii* has proven to be an efficient, albeit tricky, method for inducing infectious APMV virion production. We also showed that at least one protein associated with APMV DNA is required for APMV production after DNA microinjection. Expertise and the repetition of multiple experiments was necessary for an efficient microinjection setup, leading, in this case, to the efficient production of viral

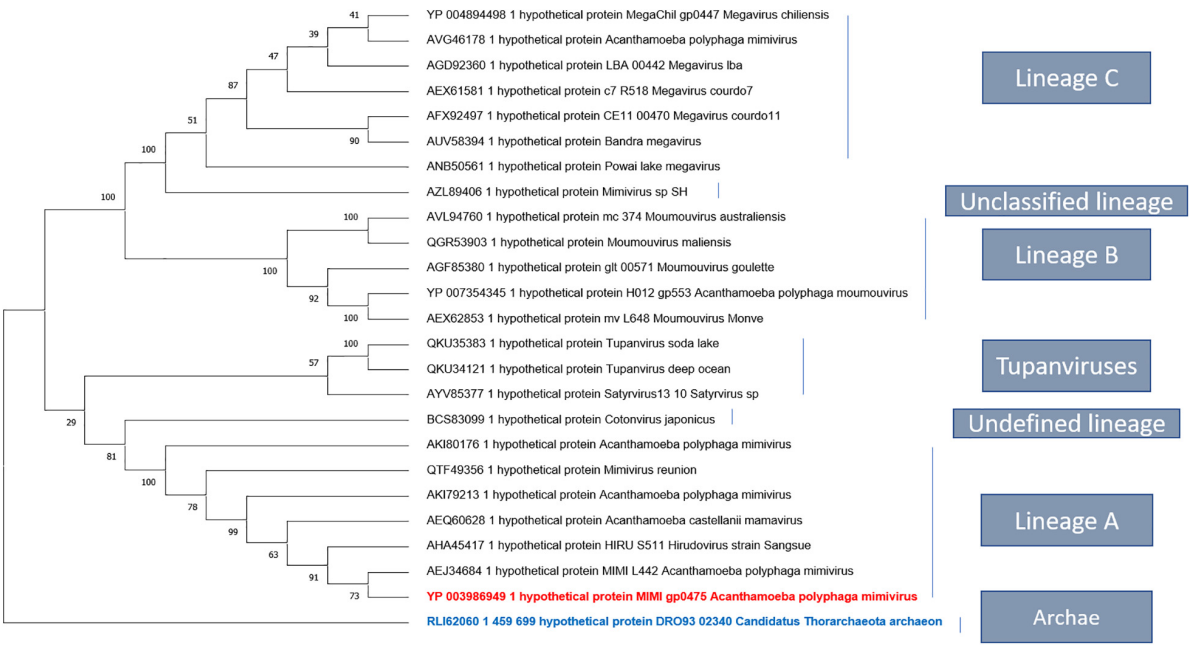


FIGURE 6 | Phylogenetic tree based on the amino acid sequences of mimivirus uncharacterized protein L442. The analysis was performed using maximum likelihood method (ML). The amino acid sequences were aligned using Muscle, and the tree was built using FastTree. Bootstraps below 50 were discarded. There were a total of 1,358 positions in the final dataset. Evolutionary analyses were conducted in MEGA X.

particles after microinjection of APMV DNA. Virions produced by microinjected amoeba were microscopically identified as mimiviruses, checked by flow cytometry, and confirmed by genome sequencing, which showed that initial mimivirus and microinjection-produced mimivirus were identical. The major drawback of this methodology was a low success rate that could be explained by various parameters. The low diameter of *A. castellanii* cells (<20 μm) and their constant morphological changes with the formation and retraction of pseudopodia (**Figure 7**) led to diverse individual cell heights and difficulties in setting Z-limits for microinjection. The amoebae also exhibit a very weak adhesion on surfaces, potentially causing detachment by capillary. We also noticed that amoebas were able to maintain microinjected samples into vacuoles and that cell membranes could be perforated, resulting in physical damage of the cells with capillary clogging. It should be noted that concentrated APMV DNA extract generated cytotoxic effects on amoeba cells, leading to the use of a diluted APMV DNA extract for further microinjection. Low concentrations of APMV DNA extracts may result in a lower performance of the microinjection. The various manipulations of DNA (extraction, double filtration, mixing the extract with the dye, loading it in femtotips, and then injecting femtotips) raise questions as to its integrity, but previous work has shown a genomic reduction of 16% for mimivirus after several passages (Boyer et al., 2011), suggesting that all genomic integrity is accessory to viral production. These technical complications cannot all be resolved, giving rise to poor performance. One of the strategies used by amoeba and avoided by microinjection is the external signaling of phagocytosis (Silva et al., 2016). This work reported that putative quorum-sensing molecules secreted by trophozoites infected with giant viruses induce the transition of neighboring cells to the cyst-resistant phase, giving them protection against giant viruses. Indeed the microinjection of an amoeba avoids the emission of the signal encystment factors, rendering the amoeba susceptible to infection. In addition to the infection of only one or some cells (less than nine cells) with the microinjection system, the absence of phagocytosis and external signaling could be the result of the slow multiplication of the virus that is followed first by the cytopathic effect of the amoeba and then by amoeba lysis with viral production.

For the first time, we have shown that the microinjection of viral DNA into its host cell leads to the production of APMV virions. This production of APMV virions was prevented by the pre-treatment of APMV DNA with proteinase K before microinjection. Protein analyses with MALDI-TOF and LC-MS were carried out, which demonstrated five interesting proteins—uncharacterized proteins L422, L724, L829, and R387 and putative GMC-type oxidoreductase R135 seem particularly interesting. Uncharacterized protein L422 was probably associated with APMV DNA and was therefore necessary for viral production after the microinjection of APMV DNA. The presence in uncharacterized protein L442 of glycosylated bonds between aa365 and aa410 (NSS, NST, and NNS) (generated with the Bruker biotools software) raises questions about other possible proteins. It would be interesting to understand whether this protein, found in a cleaved state, is in its native state or why it is in this state. However, this work on protein analysis needs further investigation. The similarity of the three-dimensional structure of this protein to a human ATP-dependent DNA helicase suggests that this protein could be involved in the access of the amoeba replication machinery to the dsDNA extracted by APMV. The most widely reported and best described DNA-related proteins in giant viruses are homologs of core histones, which are found in the viral particles of Marseilleviridae (Erives, 2017), *Acanthamoeba castellanii* medusavirus (Yoshikawa et al., 2019), and Clandestinovirus (unpublished data). In our work, these histones were not detected on viral DNA.

Previous studies have already identified certain proteins found in our work. On the one hand, Boyer et al. (2011) identified two

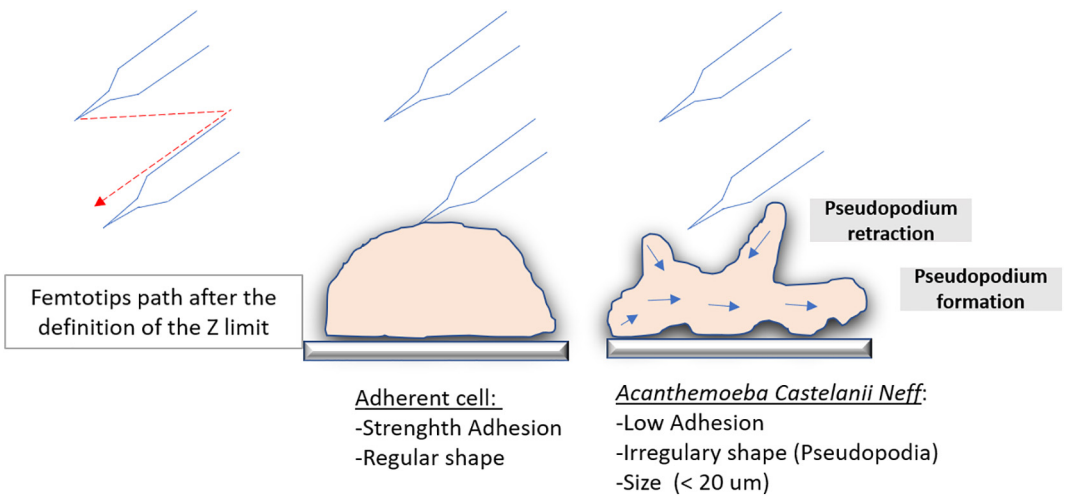


FIGURE 7 | Microinjection difference on highly adherent cells and on amoeba. Schematic illustration of the trajectory comparison of the femtotips on a highly adherent cell and on an amoeba that is a less adherent cell during the initiation of the microinjection.

glycosylated proteins (L829 and R135) representing an antigenic part of Mimivirus fibrils. Concerning these same fibrils, Araújo et al. (2015) have described an original mechanism corresponding to their strong adhesion which is mediated by glycans, specifically mannose and N-acetylglucosamine (a monomer of chitin and peptidoglycan), allowing their attachment to different organisms, especially to amoebae and virophages. Without this adhesion capacity of the fibrils, mimiviruses will not interact with virophages and amoebae (Boyer et al., 2011). In addition, mimivirus protein R135, associated to fibrils, was found in the protein panel of the virophage Sputnik (La Scola et al., 2008).

On the other hand, Bekliz et al. (2018) demonstrated that silencing of the R458 gene, encoding the R458 protein predicted for the initiation of translation, induces a deregulation of the expression of 32 proteins. Among the five proteins identified in this work, four of them—uncharacterized proteins L442, L724, and L829 and putative GMC oxidoreductase R135—are included in this set of deregulated proteins. Indeed these protein deregulations are generally associated with viral particle structure, transcription machinery, oxidative pathways, protein/lipid modifications, and DNA topology and repair. L442 with unknown function was found in both cases, with nine spots downregulated against four spots upregulated. L724, with unknown function, and R135, involved in oxidative pathway, were found in the upregulated spots. L829, with unknown function, was found in the downregulated spots (Bekliz et al., 2018).

CONCLUSION AND PERSPECTIVE

Inoculation of APMV DNA extract into *A. castellanii* by microinjection has proven to be an efficient, albeit tricky, method of inducing infectious APMV virion production. Giant virus virions not only encompass RNA, dsDNA, and proteins but also DNA-associated proteins, whose role is mandatory for infecting cells after DNA microinjection. Uncharacterized proteins L422, L724, L829, and R387 and putative GMC-type oxidoreductase R135 may be involved in APMV dsDNA availability to process through the amoeba replication machinery, and it would be interesting to express them in order to study its structure. This innovative methodology with APMV DNA extract microinjection into amoeba may be more broadly applied to other giant viruses or even non-giant classical DNA viruses and thus represents a powerful tool in the field of virology. This microinjection method should make it possible to further analyze the relationships between giant viruses, amoebae, and especially virophages. Due to the limited size of their genome, indeed virophages are more easily genetically modified. In fact, it becomes potentially possible by using this microinjection technique to inject the DNA of the susceptible giant virus and that of the modified virophage at one time. It thus becomes possible to study the effect of each virophage gene (by KO, for example) or to test the effect of these modifications on the capacity of the MIMIVIRE system to destroy the virophage. Microinjected amoebae can further be individually cloned by combining single-cell microaspirations (Sahmi-bounsiar et al., 2019).

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding author/s.

AUTHOR CONTRIBUTIONS

DS-B performed the microinjection experiments and wrote the manuscript. J-PB performed the development of microinjection experiments and wrote the manuscript. PD performed the proteomics experiments and analyzed the data. EC and SA wrote the manuscript. BL designed and supervised the study and wrote the manuscript. All authors read the final version of the manuscript.

FUNDING

This work was supported by the French Government under the "Investments for the Future" program managed by the National Agency for Research (ANR), Méditerranée-Infection 10-IAHU.

ACKNOWLEDGMENTS

We sincerely thank Takashi Irie, Kyoko Imai, Shigeki Matsubara, Taku Sakazume, Toshihide Agemura, Yusuke Ominami, Akiko Hisada, and the Hitachi High-Tech Corporation team in Japan for the collaborative study conducted together with IHU Méditerranée Infection, and the installation of TM4000 Plus and SU5000 microscopes at the IHU Méditerranée Infection facility.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2021.677847/full#supplementary-material

Supplementary Video 1 | Amoeba viability and motility 24 h after microinjection.

REFERENCES

Araú, R., Rodrigues, L., Natas, J., and Abrahã, S. (2016). Giants among larges: how gigantism impacts giant virus entry into amoebae. *Curr. Opin. Microbiol.* 31, 88–93. doi: 10.1016/j.mib.2016.03.009

Araújo, R., Rodrigues, L., Silva, S., Dornas, P., Oliveira, B., De, et al. (2015). Mimivirus Fibrils Are Important for Viral Attachment to the Microbial World by a Diverse Glycoside Interaction Repertoire. *J. Virol.* 89, 11812–11819. doi: 10.1128/JVI.01976-15.Editor

Bekliz, M., Azza, S., Decloquement, P., Raoult, D., and La Scola, B. (2018). Experimental Analysis of Mimivirus Translation Initiation Factor 4a Reveals

Its Importance in Viral Protein Translation during Infection of Acanthamoeba polyphaga. *J. Virol.* 92, 1–13.

Boyer, M., Azza, S., Barrassi, L., Klose, T., Campocasso, A., Pagnier, I., et al. (2011). Mimivirus shows dramatic genome reduction after intraamoebal culture. *Proc. Natl. Acad. Sci. U. S. A.* 108, 10296–10301. doi: 10.1073/pnas.1114908108

Byers, T. J. (1979). Growth, reproduction and differentiation in Acantha- moeba. *Int. Rev. Cytol.* 61, 283–338. doi: 10.1016/s0074-7696(08)62000-8

Dean, D. A., and Gasiorowski, J. Z. (2010). Microinjecting Cells Using a Pulsed-Flow Microinjection System. *Cold Spring Harb. Protoc.* 2011:rot5589. doi: 10.1101/pdb.prot5589

Heukeshoven, J., and Dernick, R. (1985). Simplified method for silver staining of proteins in polyacrylamide gels and the mechanism of silver staining. *Electrophoresis* 6, 103–112. doi: 10.1002/elps.1150060302

Eisenstark, A. (1965). Transduction of *Escherichia coli* Genetic Material by Phage p22 in *Salmonella* Typhimurium × E. coli Hybrids. *Proc. Natl. Acad. U. S. A.* 54, 1557–1560. doi: 10.1073/pnas.54.6.1557

Erives, A. J. (2017). Phylogenetic analysis of the core histone doublet and DNA topo II genes of Marseilleviridae: evidence of proto-eukaryotic provenance. *Epigenet. Chromatin.* 10:55. doi: 10.1186/s13072-017-0162-0

Fouque, E., Trouilhé, M., Thomas, V., Hartemann, P., Rodier, M., and Héchard, Y. (2012). Cellular, biochemical, and molecular changes during encystment of free-living amoebae. *Eukaryot Cell* 11, 382–387. doi: 10.1128/EC.05301-11

Guglielmini, J., Woo, A. C., Krupovic, M., Forterre, P., and Gaia, M. (2019). Diversification of giant and large eukaryotic dsDNA viruses predated the origin of modern eukaryotes. *Proc. Natl. Acad. U. S. A.* 116, 19585–19592. doi: 10.1073/pnas.1912006116

Hasni, I., Chelkha, N., Baptiste, E., Rayane, M., and Lachuer, J. (2019). Investigation of potential pathogenicity of Willaertia magna by investigating the transfer of bacteria pathogenicity genes into its genome. *Sci. Rep.* 9:18318. doi: 10.1038/s41598-019-54580-6

Keith, C., DiPaola, M., Maxfield, F. R., and Shelanski, M. L. (1983). Microinjection of Ca++-calmodulin causes a localized depolymerization of microtubules. *J. Cell Biol.* 97, 1918–1924. doi: 10.1083/jcb.97.6.1918

Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N., and Sternberg, M. J. E. (2015). The Phyre2 web portal for protein modeling, prediction and analysis. *Nat. Protoc.* 10, 845–858. doi: 10.1038/nprot.2015.053

Kim, T. K., and Eberwine, J. H. (2010). Mammalian cell transfection: the present and the future. *Anal. Bioanal. Chem.* 397, 3173–3178. doi: 10.1007/s00216-010-3821-6

Król-turmińska, K., and Olender, A. (2017). Human infections caused by free-living amoebae. *Ann. Agric. Environ. Med.* 24, 254–260. doi: 10.5604/12321966.1233568

La Scola, B., Audic, S., Robert, C., Jungang, L., de Lamballerie, X., Drancourt, M., et al. (2003). A giant virus in amoebae. *Science* 299:2033. doi: 10.1126/science.1081867

La Scola, B., Desnues, C., Pagnier, I., Robert, C., Barrassi, L., Fournous, G., et al. (2008). The virophage as a unique parasite of the giant mimivirus. *Nature* 455, 100–104. doi: 10.1038/nature07218

Lopata, A., Johnston, I. W. H., Hoult, I. J., and Speirs, A. I. (1980). Pregnancy Following Intrauterine Implantation Of An Embryo. *Fertil. Steril.* 33, 117–120. doi: 10.1016/S0015-0282(16)44529-2

Merril, C. R., Dunau, M. L., and Goldman, D. (1981). Silver Stain for Polypeptides in Polyacrylamide Gels. *Anal. Biochem.* 110, 201–207. doi: 10.1016/0003-2697(81)90136-6

Mougari, S., Abrahao, J., Oliveira, G. P., Khalil, J. Y. B., and La Scola, B. (2019). Role of the R349 Gene and Its Repeats in the MIMIVIRE Defense System. *Front. Microbiol.* 10:1147. doi: 10.3389/fmicb.2019.01147

Page, A. J., Cummins, C. A., Hunt, M., Wong, V. K., Reuter, S., Holden, M. T. G., et al. (2015). Roary: Rapid large-scale prokaryote pan genome analysis. *Bioinformatics* 31, 3691–3693. doi: 10.1093/bioinformatics/btv421

Page, A. J., Taylor, B., Delaney, A. J., Soares, J., Seemann, T., Keane, J. A., et al. (2016). SNP-sites: rapid efficient extraction of SNPs from multi-FASTA alignments. *Microb. Genom.* 2:e000056. doi: 10.1099/mgen.0.000056

Peng, Z., Omaruddin, R., and Bateman, E. (2005). Stable transfection of Acanthamoeba castellanii. *Biochim. Biophys. Acta* 1743, 93–100. doi: 10.1016/j.bbamcr.2004.08.014

Raoult, D., and Boyer, M. (2010). Amoebae as genitors and reservoirs of giant viruses. *Intervirology* 53, 321–329. doi: 10.1159/000312917

Read, G. A., and Kabana, E. M. (1980). Structural Responses Of Amoebae To The Injection Of Heterologous. *J. Cell Sci.* 14, 1–14. doi: 10.1242/jcs.45.1.1

Rolland, C., Andreani, J., Louazani, A. C., Aherfi, S., Francis, R., Rodrigues, R., et al. (2019). Discovery and Further Studies on Giant Viruses at the IHU Mediterranee Infection That Modified the Perception of the Virosphere. *Viruses* 11:312. doi: 10.3390/v11040312

Sahmi-bounsiar, D., Victor, P., Boratto, D. M., Oliveira, G. P., Yaacoub, J., Khalil, B., et al. (2019). Single Cell Micro-aspiration as an Alternative Strategy to Fluorescence- activated Cell Sorting for Giant Virus Mixture Separation. *J. Vis. Exp.* 27:152. doi: 10.3791/60148

La Scola, B. (2014). Looking at protists as a source of pathogenic viruses. *Microb. Pathog.* 77, 131–135. doi: 10.1016/j.micpath.2014.09.005

La Scola, B, Mezi, L., and Weiller, P. J. (2001). Isolation of Legionella anisa Using an Amoebic Coculture Procedure. *J. Clin. Microbiol.* 39, 365–366. doi: 10.1128/JCM.39.1.365

Seemann, T. (2014). Genome analysis Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 30, 2068–2069. doi: 10.1093/bioinformatics/btu153

Silva, L. K., dos, S., Boratto, P. V. M., La Scola, B., Bonjardim, C. A., and Abrahão, J. S. (2016). Acanthamoeba and mimivirus interactions: the role of amoebal encystment and the expansion of the 'Cheshire Cat' theory. *Curr. Opin. Microbiol.* 31, 9–15. doi: 10.1016/j.mib.2016.01.003

Sinard, J. H., and Pollard, T. D. (1989). Microinjection into Acanthamoeba castellanii of monoclonal antibodies to myosin-II slows but does not stop cell locomotion. *Cell Motil. Cytoskeleton* 12, 42–52. doi: 10.1002/cm.970120106

Suzan-Monti, M., La Scola, B., Barrassi, L., Espinosa, L., and Raoult, D. (2007). Ultrastructural characterization of the giant volcano-like virus factory of Acanthamoeba polyphaga Mimivirus. *PLoS One* 2:e0328. doi: 10.1371/journal.pone.0000328

Switzer, C. (1979). Sensitive Silver Stain for Detecting Proteins Peptides in Polyacrylamide Gels. *Anal. Biochem.* 237, 231–237. doi: 10.1016/0003-2697(79)90732-2

Wang, W., Sun, Y., Zhang, M., Anderson, R., Langille, L., and Chan, W. (2008). A system for high-speed microinjection of adherent cells. *Rev. Sci. Instrum.* 79:104302. doi: 10.1063/1.3006000

Weisman, R. A., and Korn, E. D. (1967). Phagocytosis of Latex Beads by Acanthamoeba. I. Biochemical Properties. *Biochemistry* 6, 485–497. doi: 10.1021/bi00854a017

Yaacoub, J., Khalil, B., Andreani, J., Scola, B., and La. (2016). ScienceDirect Updating strategies for isolating and discovering giant viruses. *Curr. Opin. Microbiol.* 31, 80–87. doi: 10.1016/j.mib.2016.03.004

Yoshikawa, G., Blanc-Mathieu, R., Song, C., Kayama, Y., Mochizuki, T., Murata, K., et al. (2019). Medusavirus, a Novel Large DNA Virus Discovered from Hot Spring Water. *J. Virol.* 93, e2130–e2118. doi: 10.1128/JVI.02130-18

Zhang, Y., and Yu, L. (2008). Single-cell microinjection technology in cell biology. *Bioessays* 30, 606–610. doi: 10.1002/bies.20759

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Microbiology

ORIGINAL RESEARCH
published: 16 July 2021
doi: 10.3389/fmicb.2021.704052

Phylogeny of the *Varidnaviria* Morphogenesis Module: Congruence and Incongruence With the Tree of Life and Viral Taxonomy

Anthony C. Woo[1,2,3], Morgan Gaia[4], Julien Guglielmini[5], Violette Da Cunha[2,3†] and Patrick Forterre[2,3]**

[1] Pôle Analyse de Données UMS 2700 2AD, Muséum National d'Histoire Naturelle, Paris, France, [2] Département de Microbiologie, Institut Pasteur, Paris, France, [3] Université Paris-Saclay, CEA, CNRS, Institute for Integrative Biology of the Cell (I2BC), Gif-sur-Yvette, France, [4] Génomique Métabolique, Génoscope, Institut François Jacob, CEA, CNRS, Univ. Évry, Université Paris-Saclay, Évry, France, [5] Hub de Bioinformatique et Biostatistique - Département Biologie Computationnelle, Institut Pasteur, Paris, France

OPEN ACCESS

Edited by:
Julien Andreani,
IHU Mediterranee Infection, France

Reviewed by:
Frank O'Neill Aylward,
Virginia Tech, United States
David D. Dunigan,
University of Nebraska–Lincoln, United States
Thomas Klose,
Purdue University, United States

***Correspondence:**
Anthony C. Woo
anthony.woo@mnhn.fr
Patrick Forterre
patrick.forterre@pasteur.fr

†Present address:
Violette Da Cunha,
Université Paris-Saclay, CEA, CNRS, Institute for Integrative Biology of the Cell (I2BC), Gif-sur-Yvette, France

Specialty section:
This article was submitted to Virology, a section of the journal Frontiers in Microbiology

Received: 01 May 2021
Accepted: 02 June 2021
Published: 16 July 2021

Citation:
Woo AC, Gaia M, Guglielmini J, Da Cunha V and Forterre P (2021) Phylogeny of the Varidnaviria Morphogenesis Module: Congruence and Incongruence With the Tree of Life and Viral Taxonomy. Front. Microbiol. 12:704052. doi: 10.3389/fmicb.2021.704052

Double-stranded DNA viruses of the realm *Varidnaviria* (formerly PRD1-adenovirus lineage) are characterized by homologous major capsid proteins (MCPs) containing one (kingdom: *Helvetiavirae*) or two β-barrel domains (kingdom: *Bamfordvirae*) known as the jelly roll folds. Most of them also share homologous packaging ATPases (pATPases). Remarkably, *Varidnaviria* infect hosts from the three domains of life, suggesting that these viruses could be very ancient and share a common ancestor. Here, we analyzed the evolutionary history of *Varidnaviria* based on single and concatenated phylogenies of their MCPs and pATPases. We excluded *Adenoviridae* from our analysis as their MCPs and pATPases are too divergent. *Sphaerolipoviridae*, the only family in the kingdom *Helvetiavirae*, exhibit a complex history: their MCPs are very divergent from those of other *Varidnaviria*, as expected, but their pATPases groups them with *Bamfordvirae*. In single and concatenated trees, *Bamfordvirae* infecting archaea were grouped with those infecting bacteria, in contradiction with the cellular tree of life, whereas those infecting eukaryotes were organized into three monophyletic groups: the *Nucleocytoviricota* phylum, formerly known as the Nucleo-Cytoplasmic Large DNA Viruses (NCLDVs), *Lavidaviridae* (virophages) and Polintoviruses. Although our analysis mostly supports the recent classification proposed by the International Committee on Taxonomy of Viruses (ICTV), it also raises questions, such as the validity of the *Adenoviridae* and *Helvetiavirae* ranking. Based on our phylogeny, we discuss current hypotheses on the origin and evolution of *Varidnaviria* and suggest new ones to reconcile the viral and cellular trees.

Keywords: evolution, dsDNA viruses, NCLDV, giant viruses, viral taxonomy

INTRODUCTION

Studying virus origin and evolution is a challenging exercise, especially when addressing early co-evolution with their cellular hosts. While cellular domains (Archaea, Bacteria, and Eukarya) have been established based on ribosomal RNA sequences and recovered later in many single universal protein trees (Woese et al., 1990), viral "realms" have been recently proposed by the

International Committee on Taxonomy of Viruses (ICTV), based on proteins involved in virion morphogenesis and/or in viral genome replication (Koonin et al., 2020). To date, only viruses from the realms *Duplodnaviria* and *Varidnaviria*, both corresponding mostly to double-stranded (ds) DNA viruses, infect hosts from the three domains of life (Koonin et al., 2020). These realms were previously recognized as lineages, based on the conservation of their major capsid proteins (MCPs). *Duplodnaviria* and *Varidnaviria* were known as the HK97 and the PRD1-adenovirus lineages, respectively (Bamford, 2003; Baker et al., 2005; Abrescia et al., 2012). The *Duplodnaviria* mostly consists of archaeal and bacterial viruses, whereas *Varidnaviria* are well represented in the virosphere associated with all three domains. This realm is thus an ideal subject to study the evolution of viruses in the context of the universal tree of life.

The *Varidnaviria* encompasses many very diverse families (hence its name, Various DNA viruses) (**Table 1**) (Koonin et al., 2019, 2020). They are all double-stranded DNA viruses, except the FLiP single-stranded DNA viruses (Laanto et al., 2017). Notably, the sizes of their virions vary from very small to the most gigantic ones among viruses. In the new ICTV taxonomy, *Varidnaviria* are divided into two kingdoms: *Bamfordvirae*, characterized by a single MCP with a double jelly roll (DJR) fold and *Helvetiavirae* characterized by two MCPs, each with a single jelly roll fold (SJR) (Koonin et al., 2020) (**Table 1**).

The kingdom *Helvetiavirae* only includes viruses infecting archaea or bacteria. All known viruses of this kingdom are closely related and have been grouped into a single family, the *Sphaerolipoviridae* (Gil-Carton et al., 2015; Demina et al., 2017). It has been suggested that *Bamfordvirae* originated from *Helvetiavirae* by ancestral gene fusion of the SJR folds of their two MCPs (Krupovič and Bamford, 2008; Krupovic et al., 2020). In contrast to *Helvetiavirae*, the kingdom *Bamfordvirae* includes many families of viruses infecting members from the three domains. In the ICTV classification, *Bamfordvirae* have been divided into two phyla, *Nucleocytoviricota*, which includes all large to giant *Varidnaviria*, and *Preplasmiviricota*, which includes all small *Bamfordvirae* (Koonin et al., 2020).

The *Nucleocytoviricota*, formerly known as the Nucleo-Cytoplasmic Large DNA Viruses (NCLDVs), only infect eukaryotes, whereas *Preplasmiviricota* again includes viruses infecting members of the three domains. *Nucleocytoviricota* have been divided into two classes, *Megaviricetes* and *Pokkesviricetes* whereas *Preplasmiviricota* have been divided between *Maveriviricetes* and *Tectiliviricetes* (**Table 1**). It has been proposed that Polintoviruses belong to *Preplasmiviricota*. These elusive viruses are related to mobile elements called Polintons that carry genes encoding the MCP and packaging ATPases (pATPases) typical of *Varidnaviria* (Krupovic et al., 2014). *Maveriviricetes* include a single family, *Lavidaviridae* (also dubbed virophages) of viruses infecting the virocells of *Mimiviridae,* whereas *Tectiliviricetes* includes again several families of viruses infecting members of the three domains. The only *Tectiliviricetes* infecting eukaryotes correspond to *Adenoviridae*; all other *Tectiliviricetes* infecting prokaryotes, either archaea or bacteria. The best-known archaeal and bacterial *Tectiliviricetes* are small viruses, such as *Tectiviridae*, *Turriviridae*, and *Corticoviridae*, exemplified by the virus PRD1 infecting *Escherichia coli*, the virus STIV (Sulfolobus Turreted Icosahedral Virus) infecting *Sulfolobus*, and the virus PM2 infecting *Pseudoalteromonas*, respectively (San Martín and van Raaij, 2018; Yutin et al., 2018). The name *Tectiliviricetes* (Tectivirid-like and the suffix viricetes for class taxa) was designed from the best-studied virus of this class, the *Tectiviridae* PRD1. Most of these viruses are also known to integrate into bacterial or archaeal genomes (Yutin et al., 2018) or exist as free plasmids corresponding to defective viruses (Gaudin et al., 2014).

Besides the few representatives of *Turriviridae*, *Tectiviridae*, and *Corticoviridae* already known, Koonin and co-workers identified in metagenome-associated genomes (MAGs) many new lineages of archaeal and bacterial *Tectiliviricetes*. They proposed their classification into six groups based on sequence similarities networks of their MCPs and detection of signature proteins specific for each group (**Table 1**) (Yutin et al., 2018). The groups PM2, STIV and PRD1 could correspond to the orders *Vinavirales*, *Belfryvirales*, and *Kalamavirales* of the ICTV classification, respectively, whereas the Odin, Bam35/Toil and FLiP groups remained unclassified (**Table 1**). *Autolykiviridae*, a family of viruses abundant in marine microbial metagenomes (Kauffman et al., 2018) was included in the PM2 group by Koonin and colleagues (Yutin et al., 2018).

The group Odin was named after an integrated element present in the MAG of an Odinarchaeon, but all other members of this group were detected in bacterial MAGs. All other *Tectiliviricetes* infecting archaea were included in the group STIV, named from the archaeovirus STIV member. The STIV group itself was divided into two subgroups based on their MCP phylogeny, one including archaeoviruses and the other bacterioviruses (Yutin et al., 2018). The four other groups defined by Koonin and colleagues (PM2, PRD1, Bam35/Toil, and FLiP) include only bacterioviruses (Yutin et al., 2018) (**Table 1**).

Several hypotheses have been proposed regarding the origin and evolution of *Varidnaviria*. Several authors suggested that *Varidnaviria* predated the Last Universal Common Ancestor (LUCA) (Bamford et al., 2005) and that the transition from *Helvetiavirae* to *Bamfordvirae* took place before LUCA (Krupovic et al., 2020). Koonin, Krupovic, and colleagues even suggested that the diversification of bacterial and archaeal *Tectiliviricetes* predated LUCA (Krupovic et al., 2020). For these authors, *Varidnaviria* infecting eukaryotes originated later from a Tectivirus infecting the bacterium at the origin of mitochondria (Krupovic and Koonin, 2015). It is worth noting that, if this scenario is correct, the new ICTV ranking of *Varidnaviria* does not follow the rules of phylogenetic systematics, sensu Hennig (1965), since both *Helvetiavirae* and *Tectiliviricetes* are paraphyletic (e.g., the last common ancestor of *Helvetiavirae* was an ancestor of *Tectiliviricetes* and the last common ancestor of *Tectiliviricetes* was also an ancestor of all *Varidnaviria* infecting eukaryotes).

To evaluate the validity of the above scenario and eventually propose new ones, as well as to test the validity of the

newly proposed ICTV classification, it is essential to determine the evolutionary history of *Varidnaviria* based on robust phylogenetic analyses. Several trees based on the structural comparison of MCPs from a set of *Bamfordvirae* representatives have been published (Ravantti et al., 2013, 2020). The evolutionary relationships among *Bamfordvirae* were also investigated using pairwise amino-acid sequence similarities networks (Sinclair et al., 2017; Yutin et al., 2018). Although such studies can provide interesting information for further analyses, they cannot completely replace sequence-based phylogenetic analyses in determining the actual evolutionary history of biological objects or organisms.

Until now, sequence-based phylogenetic analyses dealing with *Varidnaviria* have only focused on subgroups of *Bamfordvirae*. Most of them have specifically addressed the evolution of *Nucleocytoviricota* and the origin of giant viruses. They have shown that gigantism most likely originated several times independently during the evolution of *Nucleocytoviricota* (Guglielmini et al., 2019; Koonin and Yutin, 2019). The ICTV classification of this phylum is congruent with phylogenetic analyses of five marker genes conserved in most *Nucleocytoviricota* (Koonin and Yutin, 2019). However, in our recent phylogenetic analysis based on 8 marker genes (Guglielmini et al., 2019), we observed some discrepancies, in particular related to the position of *Asfarviridae* and related viruses. A global phylogeny of prokaryotic *Tectiliviricetes* based on their MCPs was also published (Kauffman et al., 2018). In this phylogeny, STIV-related archaeoviruses branched within *Tectiliviricetes* infecting bacteria, suggesting a virus transfer from Bacteria to Archaea. This is surprising since another recent phylogeny focusing on STIV-related archaeoviruses suggested that these viruses predated the last archaeal common ancestor (LACA) (Krupovic et al., 2019). The latter phylogeny was based on the concatenation of their MCP and their pATPases. Besides a common MCP, most members of *Varidnaviria* share indeed a homologous pATPase of the FtsK/HerA superfamily P-loop ATPases (**Table 1**) (Abrescia et al., 2012; Yutin et al., 2018). Concatenation of the MCP and pATPase proved also useful in rooting the phylogenetic tree of *Nucleocytoviricota* with the Polintoviruses as an outgroup (Guglielmini et al., 2019), questioning the possibility to extrapolate such an approach to the whole *Varidnaviria* realm.

Here, we have revisited the distribution and structural similarities of MCPs and pATPases among viruses of the realm *Varidnaviria* to identify the viruses that could be relevant to the evolutionary history of their morphogenesis module. We found that these two proteins can be used as phylogenetic markers for most members of the kingdom *Bamfordvirae*. We excluded *Adenoviridae*, Odin and FLiP groups from our analysis because they lack homologous pATPase and/or the sequences of their MCP are too divergent. The concatenation of the MCP and pATPase sequences produces a rather robust viral phylogeny that can be useful to infer the history of the *Bamfordvirae* morphogenesis module. *Helvetiavirae* were excluded from our concatenation because their MCPs were too divergent, but we could include them in our single pATPase tree. Our analyses validate some of the recent taxonomic proposals but challenge others, such as the ranking of *Sphaerolipoviridae* in a specific kingdom or the grouping

TABLE 1 | Viruses of the realm *Varidnaviria*.

	MCP	pATPase	Domain	Kingdom	Phylum	Class	Order
Poxviridae	DJR	FtsK/HerA family	Eukarya	*Bamfordvirae*	*Nucleocytoviricota*	*Pokkesviricetes*	*Chitovirales*
Asfarviridae	DJR	FtsK/HerA family	Eukarya	*Bamfordvirae*	*Nucleocytoviricota*	*Pokkesviricetes*	*Asfuvirales*
Phycodnaviridae	DJR	FtsK/HerA family	Eukarya	*Bamfordvirae*	*Nucleocytoviricota*	*Megaviricetes*	*Algavirales*
Mollivirus	DJR	FtsK/HerA family	Eukarya	*Bamfordvirae*	*Nucleocytoviricota*	*Megaviricetes*	*Algavirales*
Mimiviridae	DJR	FtsK/HerA family	Eukarya	*Bamfordvirae*	*Nucleocytoviricota*	*Megaviricetes*	*Imitervirales*
Ascoviridae	DJR	FtsK/HerA family	Eukarya	*Bamfordvirae*	*Nucleocytoviricota*	*Megaviricetes*	*Pimascovirales*
Iridoviridae	DJR	FtsK/HerA family	Eukarya	*Bamfordvirae*	*Nucleocytoviricota*	*Megaviricetes*	*Pimascovirales*
Marseilleviridae	DJR	FtsK/HerA family	Eukarya	*Bamfordvirae*	*Nucleocytoviricota*	*Megaviricetes*	*Pimascovirales*
Polintoviruses	DJR	FtsK/HerA family	Eukarya	*Bamfordvirae*	*Preplasmiviricota*	/	/
Lavidaviridae	DJR	FtsK/HerA family	Eukarya	*Bamfordvirae*	*Preplasmiviricota*	*Maveriviricetes*	*Priklausovirales*
Adenoviridae	DJR	ABC family	Eukarya	*Bamfordvirae*	*Preplasmiviricota*	*Tectiliviricetes*	*Rowavirales*
STIV (*Turriviridae*)	DJR	FtsK/HerA family	Archaea/Bacteria	*Bamfordvirae*	*Preplasmiviricota*	*Tectiliviricetes*	*Belfryvirales*
Bam35	DJR	FtsK/HerA family	Bacteria	/	/	/	/
PRD1 *(Tectiviridae)*	DJR	FtsK/HerA family	Bacteria	*Bamfordvirae*	*Preplasmiviricota*	*Tectiliviricetes*	*Kalamavirales*
Toil	DJR	FtsK/HerA family	Bacteria	/	/	/	/
PM2 (*Corticoviridae*)	DJR	FtsK/HerA family	Bacteria	*Bamfordvirae*	*Preplasmiviricota*	*Tectiliviricetes*	*Vinavirales*
FLiP	DJR	/	Bacteria	/	/	/	/
Odin	DJR	/	Bacteria*	/	/	/	/
Sphaerolipoviridae	SJR	FtsK/HerA family	Bacteria/Archaea	*Helvetiavirae*	*Dividoviricota*	*Laserviricetes*	*Halopanivirales*

*This table summarizes the shared genes between different families of this lineage. DJR, double jelly roll; SJR: single jelly roll; NCLDV, Nucleo-Cytoplasmic Large DNA Viruses; STIV, Sulfolobus turreted icosahedral virus; FLiP, Flavobacterium-infecting bacteriophage; MCP, major capsid protein; pATPase, packaging ATPase. *The group Odin includes only bacterial MAG, except for one MAG of Odinarchaeota.*

of *Lavidaviridae*, Polintoviruses, and *Adenoviridae* within *Tectiliviricetes*. We confirm that *Tectiliviricetes* infecting archaea are closely related to those infecting bacteria, in contradiction with the topology of the universal cellular tree. We discuss the main hypothesis that has been previously proposed to describe the evolution of *Varidnaviria* in the light of our findings and explore alternative scenarios that could explain the discrepancy between the viral tree of *Bamfordvirae* and the universal cellular tree of life.

MATERIALS AND METHODS

Selection of MCP/pATPase Sequences

Representative MCP/pATPase sequences from different groups of *Preplasmiviricota* were used as queries for PSI-BLAST (Altschul et al., 1997) searches against the GenBank non-redundant protein sequence database (nr). The query sequences are listed below:

Group	Name	MCP	pATPase
Lavidaviridae	Sputnik virophage	YP_002122381	YP_002122364
Adenoviridae	Frog adenovirus 1	NP_062443	NP_062434
STIV	Sulfolobus turreted icosahedral virus 1	YP_025022	YP_025021
Bam35	Bacillus phage Bam35c	NP_943764	NP_943760
Tectiviridae	Enterobacteria phage PRD1	NP_040692	NP_040689
Toil	Rhodococcus phage Toil	ARK07697	ARK07695
PM2	Pseudoaltero monas phage PM2	NP_049903	NP_049900
FLiP	Flavobacterium phage FLiP	ASQ41214	-

The MCP/pATPase sequences of the *Nucleocytoviricota* were retrieved from a previous study that we conducted (Guglielmini et al., 2019). The Polintoviruses sequences were gathered from the Repbase collection (Jurka et al., 2005)[1]: Polinto-2_NV, Polinto-1_DY, Polinto-1_TC, Polinto-1_SP, Polinto-2_SP, Polinto-2_DR, and Polinto-1_DR. Finally, the sequences from the SJR group were recovered based on previously identified sequences (Demina et al., 2017).

Putative MCP/pATPase sequences were aligned with the query sequences for the examination of the conserved structural elements using MAFFT (Katoh and Standley, 2013). Prediction of the secondary structure was performed using Phyre2 (Kelley et al., 2015) and the predicted protein structures were visualized using UCSF Chimera (Pettersen et al., 2004). The sequences used in this study are shown in the **Supplementary File**. After removing sequences with no significant matches or low confidence levels, we obtained two different datasets of 145 and 128 sequences for the MCP and pATPase respectively.

Network Analysis

After performing the structural protein prediction analysis, all-against-all blastp analyses were performed on the refined pATPase datasets. The all-against-all blastP results were grouped using the SiLiX (for *SIngle LInkage Clustering of Sequences*) package v1.2.8[2] (Miele et al., 2011). This approach for the clustering of homologous sequences is based on single transitive links with alignment coverage constraints. The pATPase sequences were clustered independently by similarity using SiLiX with the expect threshold of 0.001 as previously used for MCP analysis (Yutin et al., 2018). The clustering results were analyzed and visualized using the igraph package of the R programming language[3].

Sequence Alignment

The alignments of the MCP sequences were performed using MAFFT v7.392 with the E-INS-i algorithm (Katoh and Standley, 2013), which can align sequences with several conserved motifs embedded in long unalignable regions, whereas pATPase sequences were aligned using MAFFT with the L-INS-i algorithm (Katoh and Standley, 2013), which can align a set of sequences containing sequences flanking around one alignable domain. Positions containing more than 30% of gaps were trimmed using goalign v0.2.8[4].

Phylogenetic Analysis

Single protein and concatenated protein phylogenies were conducted within the maximum likelihood (ML) framework using IQ-TREE v1.6.3 (Nguyen et al., 2015). We first performed a model test with the Bayesian Information Criterion (BIC) by including protein mixture models (Kalyaanamoorthy et al., 2017). For mixture model analyses, we used the PMSF models (Wang et al., 2018). Bootstrap trees with 1,000 replicates were generated using IQ-TREE with the same parameters as the best-known likelihood tree search. Non-parametric classical bootstrap values, as well as transfer bootstrap expectation (TBE) values (Lemoine et al., 2018) were computed using the software gotree v0.3.0[5].

Visualization

The phylogenetic trees were visualized with FigTree v1.4.3[6] and iTOL (Letunic and Bork, 2007).

[1]http://www.girinst.org/Repbase_Update.html

[2]http://lbbe.univ-lyon1.fr/SiLiX

[3]https://igraph.org/

[4]https://github.com/evolbioinfo/goalign

[5]https://github.com/evolbioinfo/gotree

[6]http://tree.bio.ed.ac.uk/software/figtree/

RESULTS

Identification of *Varidnaviria* MCPs and pATPases Suitable for Phylogenetic Analysis

We retrieved MCP and pATPase sequences using PSI-BLAST searches against the NCBI non-redundant protein sequence database (nr) and added sequences recovered from proviruses (section "Materials and Methods"). We could identify both MCP and pATPase for most *Varidnaviria* (**Table 1**) with some exceptions. In particular, we could not detect putative pATPase in members of the Odin and FLiP groups, as was previously observed (Yutin et al., 2018). To validate the identified MCP and pATPase sequences, we generated protein models for all selected sequences and compared these predicted structures to the PDB database using Phyre2 (**Supplementary Figure 1** and **Supplementary Data File 1**; Kelley et al., 2015). The *Varidnaviria* MCPs associated with groups and families previously described indeed matched their corresponding structures in the public databases, except for the putative MCPs identified from the Odin group, which was thus excluded from further analysis. The MCP from *Adenoviridae* was unique in exhibiting several additional structural elements (**Supplementary Figure 1**). We also confirmed that MCPs with SJR fold from *Sphaerolipoviridae* (*Helvetiavirae*) were very divergent from those of *Bamfordvirae*.

All *Varidnaviria* pATPases, except those of *Adenoviridae*, share similar predicted structures (**Supplementary Figure 2** and **Supplementary Data File 1**) and clustered together in an amino acid sequence similarity network (**Supplementary Figure 3**). Previous observations based on amino-acid signatures and secondary structure predictions have indeed concluded that *Adenoviridae* pATPases were not specifically related to other *Varidnaviria* pATPases but to ATPases of the ABC superfamily (**Supplementary Data File 1**), indicating an exchange of pATPases during the evolution of *Adenoviridae* (Burroughs et al., 2007). Surprisingly, unlike the situation with MCP, the pATPases of *Sphaerolipoviridae* were structurally similar to those of *Bamfordvirae* (**Supplementary Figure 2**) and clustered together with *Tectiliviricetes* in our amino acid sequence similarity network (**Supplementary Figure 3**). Another peculiarity of the pATPase network was that *Lavidaviridae* exhibited more connections with *Tectiliviricetes* than with Polintoviruses and *Nucleocytoviricota*. Our results thus suggested that, besides the MCP, the pATPase gene could be an interesting marker for delineating the phylogeny of most *Varidnaviria* and could provide interesting information on the origin and evolution of *Sphaerolipoviridae*.

Comparison of Single Trees Suggests Congruent Evolution of the MCPs and pATPases

To facilitate the comparison of the MCP and pATPase trees, we removed the taxa that were not present in both datasets, *Helvetiavirae*, FLiP, and Odin, as well as *Adenoviridae* whose MCP and pATPase could not be aligned with those of other *Varidnaviria* (**Table 1**). Phylogenetic analyses were first performed separately on the two proteins within the ML framework (section "Materials and Methods"). We thus obtained the first sequence-based phylogenies covering most groups of *Bamfordvirae* (**Figure 1** and **Supplementary Figures 4**, **5**). The MCP and pATPase trees exhibited noticeable congruence confirming that they belong to the same module (Iranzo et al., 2016b) (**Figure 1**). Notably, *Bamfordvirae* infecting prokaryotes (*Tectiliviricetes*) and those infecting eukaryotes formed two distinct clusters with good TBE support values. The only large bipartition in common between the two single-protein trees while maintaining most known large monophyletic groups corresponded to the separation between *Tectiliviricetes* and eukaryotes-infecting *Bamfordvirae*. We thus decided to root the trees between these two clusters (**Figure 1**) although there are other possibilities (see section "Discussion" and **Figure 2**).

We recovered in both trees the monophyly of most previously defined groups and/or families of eukaryoviruses. Noticeably, the *Nucleocytoviricota* phylum was not monophyletic because of the variable positions of *Asfarviridae* and *Poxviridae*. We also recovered the monophyly of the previously defined groups of *Tectiliviricetes*, except for the STIV group that was paraphyletic in both trees and the PM2 group that was paraphyletic in the pATPase tree. STIV infecting archaea and STIV infecting bacteria were monophyletic in the MCP tree but the latter were sister group to the PM2 group. In the pATPase tree, some STIV infecting archaea were mixed with the PM2 group. Noticeably, the archaeoviruses STIV branched within bacterioviruses in both cases, as previously observed (Kauffman et al., 2018). Other similarities between the MCP and pATPase trees were the grouping of the bacterioviruses STIV with the PM2 group and the proximity of the Cellulomonas, Bam35, *Tectiviridae* in both trees (the three of them forming a single clade together with the Toil group in the pATPase tree).

The small differences observed between the MCP and the pATPase trees could be due to conflicting phylogenetic signals due to lateral gene transfer and/or to the low resolution in some part of these trees. The rather good congruence between the two trees thus suggested that concatenation of the MCP and pATPase sequences could be used to obtain a more reliable phylogeny of the *Bamfordvirae* virion morphogenesis module.

Concatenation of the MCP and the pATPase Produces a Robust Tree of *Bamfordvirae*

In the concatenated tree, the number of well-supported branches increased (**Figure 3A** and **Supplementary Figure 6**). In particular, we recovered both strong bootstraps (0.95) and TBE support (0.99) for the bipartition between *Bamfordvirae* infecting eukaryotes and those infecting prokaryotes (**Figure 3A** and **Supplementary Figure 6**) although TBE supports for the grouping of *Tectiliviricetes* with Polintons (0.91) or with Polintons and *Lavidaviridae* (0.88) were also rather

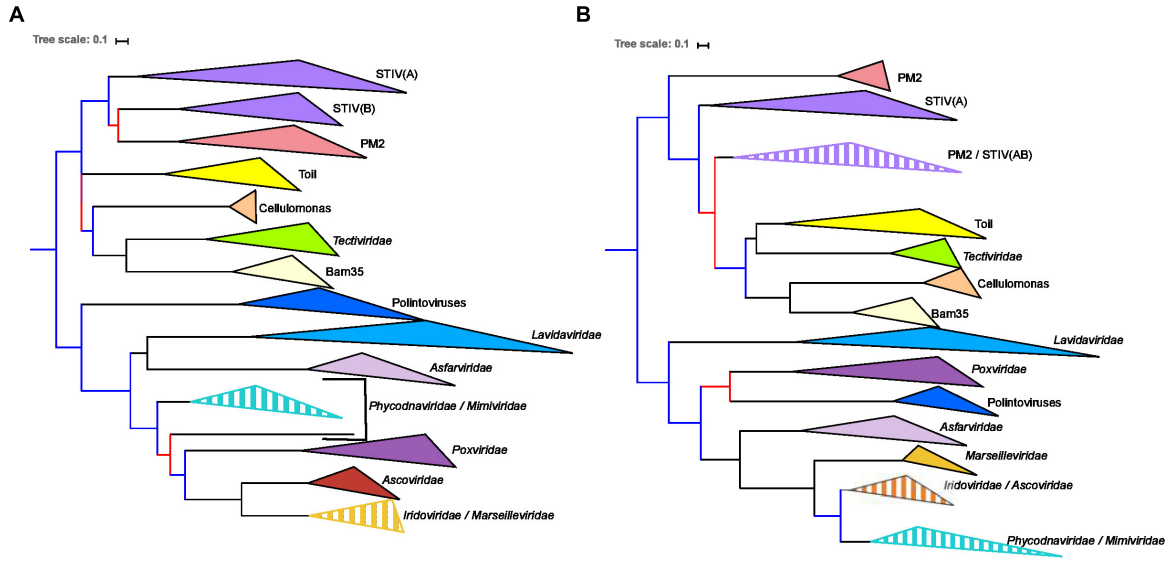


FIGURE 1 | Single-protein trees of the two hallmark proteins of the viruses from the *Varidnaviria* realm, excluding *Adenoviridae, Sphaerolipoviridae*. Phylogenetic trees of **(A)** the major capsid protein (MCP) and **(B)** packaging ATPase (pATPase). The root of the phylogenetic tree was between the prokaryotic and eukaryotic members. The scale-bar indicates the average number of substitutions per site. The best-fit model for the MCP tree was LG + F + R4, which was chosen according to Bayesian Information Criterion (BIC) and the alignment has 103 sequences with 237 positions. The best-fit model for the pATPase tree was LG + R6, which was chosen according to BIC and the alignment has 103 sequences with 171 positions. More detailed versions of the trees are shown in **Supplementary Figures 4**, **5**. Branches in black indicate both classical bootstrap and transfer bootstrap expectation (TBE) support values using 1,000 replicates are above 70%. Branches in blue indicate only one of the two support values is above 70% whereas branches in red indicate both support values are below 70%.

high. We recovered most clades suggested by the ICTV classification, except for the *Pokkesviricetes* (**Figure 3B**). The ICTV classification would have suggested rooting the tree between *Lavidaviridae* and *Poxviridae* (**Figure 3B**) (see the discussion for the different possible positions of the root). In contrast to the results obtained with the single-protein trees, *Nucleocytoviricota*, including *Poxviridae,* were monophyletic in the concatenated tree. When rooting between the eukaryoviruses and bacterioviruses, the Polintoviruses were basal to all eukaryotic groups. The *Poxviridae* were the first branching family of *Nucleocytoviricota*, followed by the *Asfarviridae*. Notably, we recovered the two major groups of *Nucleocytoviricota* that we previously identified based on 8 core genes (Guglielmini et al., 2019), the MAPI (*Marseilleviridae*, *Ascoviridae,* Pitho-like viruses, and *Iridoviridae*), which corresponds to the order "*Pimascovirales*" and the PAM [*Phycodnaviridae* (*Algavirales*), *Asfarviridae* (*Pokkesviricetes*), and Megavirales (*Imitervirales*)] except that the *Asfarviridae* were a sister group to these two superclades instead to be part of the PAM.

Interestingly, whereas the STIV group remained paraphyletic in the concatenated trees, the STIV viruses infecting archaea formed a monophyletic group (hereinafter called archaeal DJR cluster) with a good TBE support (0.88) (**Figure 3** and **Supplementary Figure 6**), as in the MCP tree (**Figure 1A**). The *Tectiliviricetes* infecting bacteria were divided into two clades. The first one (hereinafter called DJR bacterial cluster I) contained the PM2 group, including *Autolykiviridae*, and bacterial members of the STIV group. The TBE support for this group was rather weak (0.70) but still significant. The second clade, (hereinafter called DJR bacterial cluster II) was strongly supported (TBE = 0.90) contained the PRD1, Cellulomonas and the Bam35/Toil groups. The DJR bacterial cluster I emerged at the base of the *Tectiliviricetes*, whereas the DJR bacterial cluster II was the sister group to the archaeal DJR cluster with low but significant support.

We previously noticed that *Poxviridae* have long branches and variable positions in single-gene trees of *Nucleocytoviricota* proteins (Guglielmini et al., 2019). In particular, they tended to attract the long branches of the *Asfarviridae* in our previous analyses, forming a clade corresponding to the recently proposed class *Pokkesviricetes*. When we removed *Poxviridae* from our dataset to prevent possible long-branch attraction, we obtained a tree in which *Nucleocytoviricota* are no more sister group to the *Lavidaviridae* (**Figure 4** and **Supplementary Figure 7**), but Polintoviruses, with good support. In that case, *Lavidaviridae* branched between *Tectiliviricetes* and all other *Varidnaviria,* as in the pATPase tree. This position could also explain their weak clustering with *Tectiliviricetes* in the pATPase amino-acid similarity network (**Supplementary Figure 3**). The monophyly of archaeal *Tectiliviricetes*, previously observed with the complete MCP tree, was even more strongly supported (TBE = 0.97) in the concatenated trees without *Poxviridae* (**Figure 4** and **Supplementary Figure 7**). In addition, we obtained with strong TBE support the monophyly of the entire STIV group, STIV archaeoviruses branching as sister group to STIV bacterioviruses. In this tree, the DJR archaeal cluster was thus included in the DJR bacterial cluster I which became paraphyletic, whereas the DJR bacterial cluster II emerged with strong TBE support at

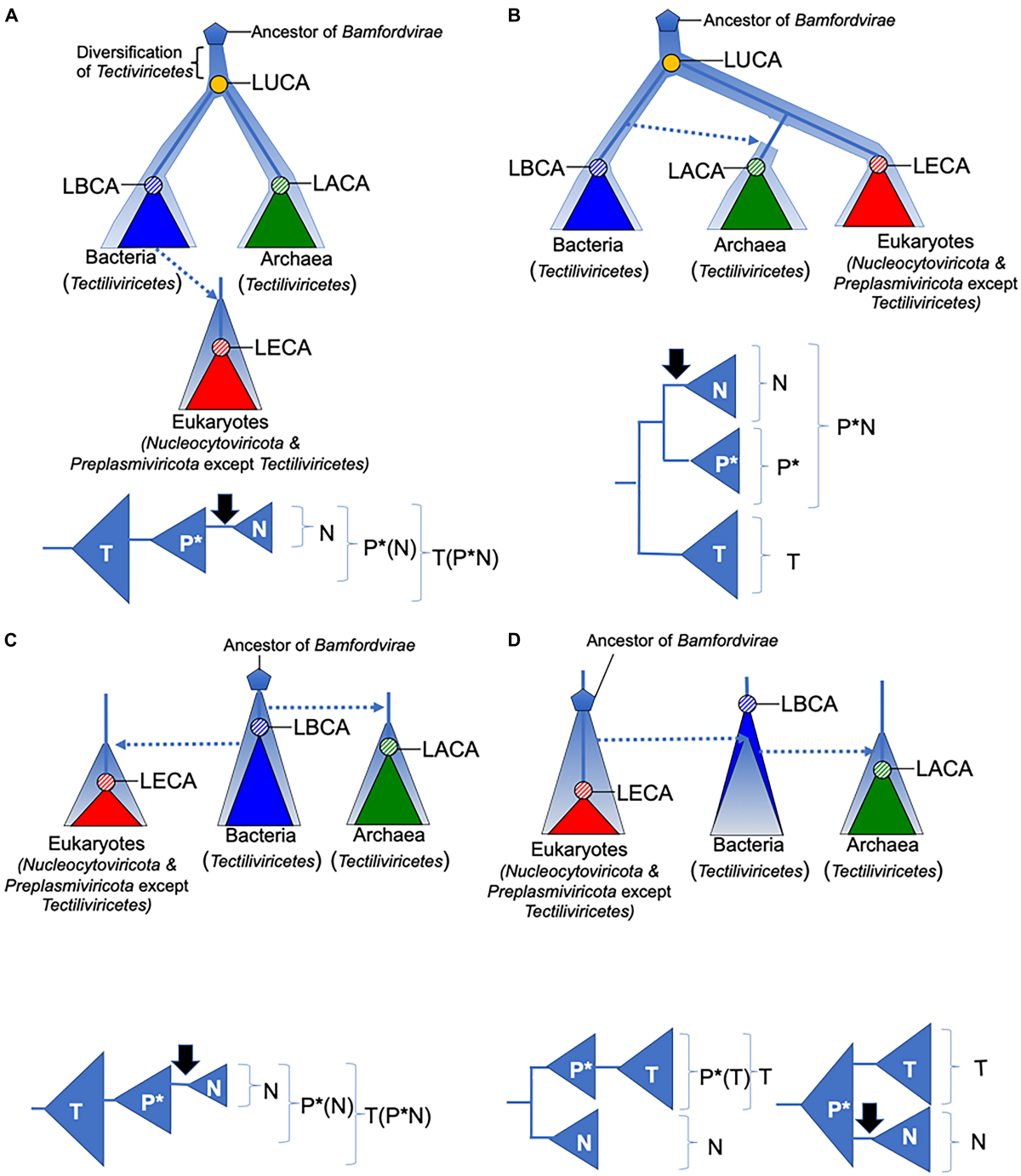


FIGURE 2 | Four scenarios for the evolution of *Bamfordvirae*. The diversification and evolution of viruses are depicted with pale blue thick lines or triangles. Lower panels depict groups that are monophyletic (in brackets) among *Tectiliviricetes* (T), *Preplasmiviricota* except *Tectiliviricetes* (P*), and *Nucleocytoviricota* (N) in the above scenarios. The black arrows indicate the introduction of DNA replication proteins related to those of *Duplodnaviria* in the ancestral lineage of *Nucleocytoviricota*. Dotted arrows indicate transfer of viruses between different cellular domains. In all scenarios, we assumed that *Bamfordvirae* were present in the last archaeal common ancestor (LACA). In panels **(A,B)**, the origin of *Bamfordvirae* (blue pentagon) predated the last universal common ancestor (LUCA). In panel **(A)**, *Tectiliviricetes* emerged first and different lineages were selected and co-evolved with the proto-bacterial and proto-archaeal lineages. In panel **(B)**, the ancestral *Varidnaviria* thriving at the time of LUCA have now disappeared, *Tectiliviricetes* emerged in the bacterial branch and were later transferred to Archaea. The ancestors of eukaryotic *Bamfordvirae* were lost in Archaea. In panels **(C,D)**, *Bamfordvirae* originated after LUCA, either in proto-bacteria **(C)** or proto-eukaryotes **(D)**. They appeared later in the two other domains by virus transfer. LUCA, last universal common ancestor; LACA, last archaeal common ancestor; LBCA, last bacterial common ancestor; LECA, last eukaryal common ancestor.

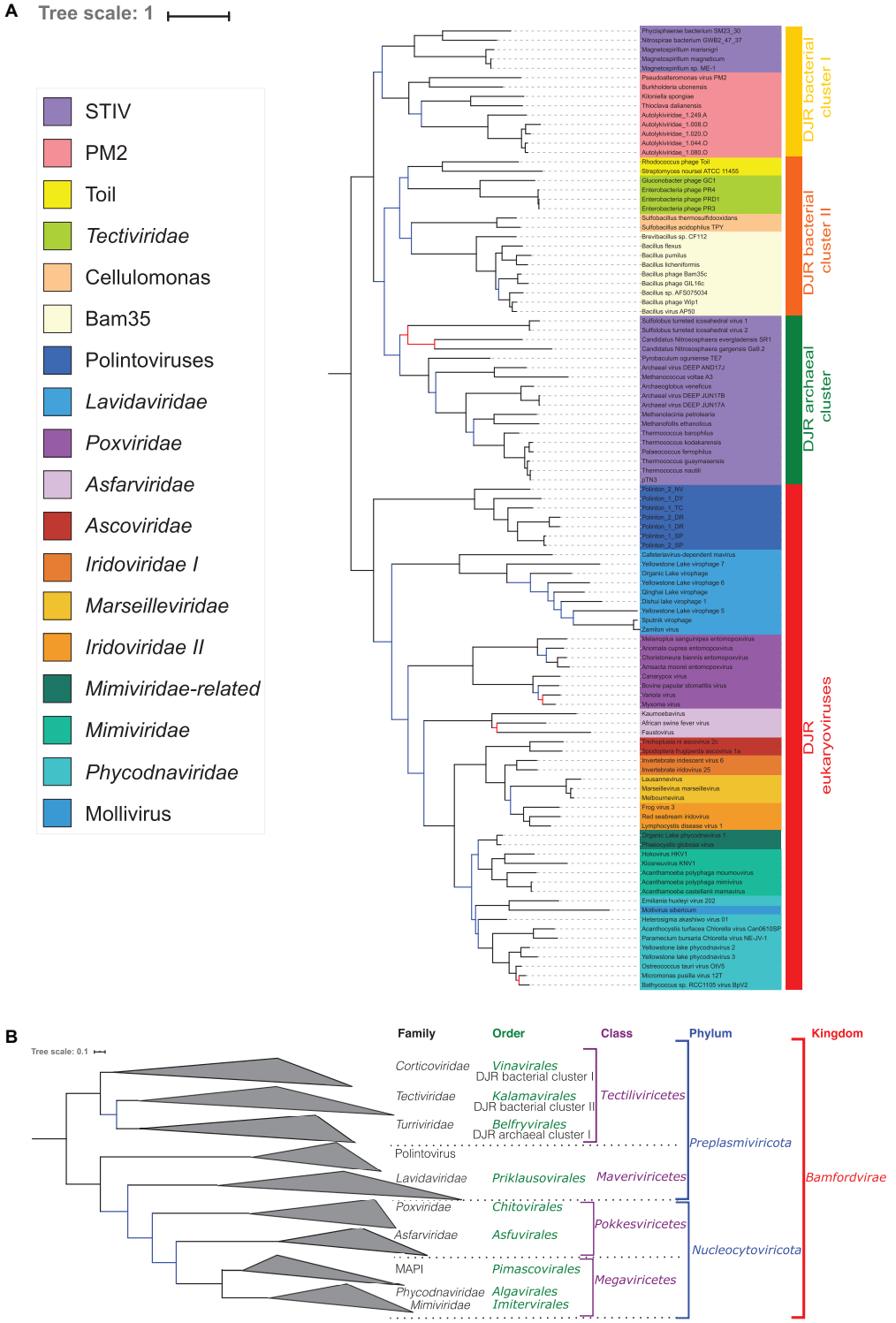


FIGURE 3 | Phylogenetic tree of the concatenated MCP and pATPase genes. **(A)** Phylogenetic tree annotated with the corresponding group. The alignment has 103 sequences with 408 positions. Phylogenetic tree was rooted between the prokaryotic and eukaryotic members. The scale-bar indicates the average number of substitutions per site. The best-fit model was LG + F + R6, which was chosen according to Bayesian Information Criterion (BIC). Branches in black indicate both classical bootstrap and transfer bootstrap expectation (TBE) support using 1,000 replicates values are above 70%. Branches in blue indicate only one of the two support values is above 70% whereas branches in red indicate both support values are below 70%. **(B)** Phylogenetic tree annotated with the ICTV taxonomy.

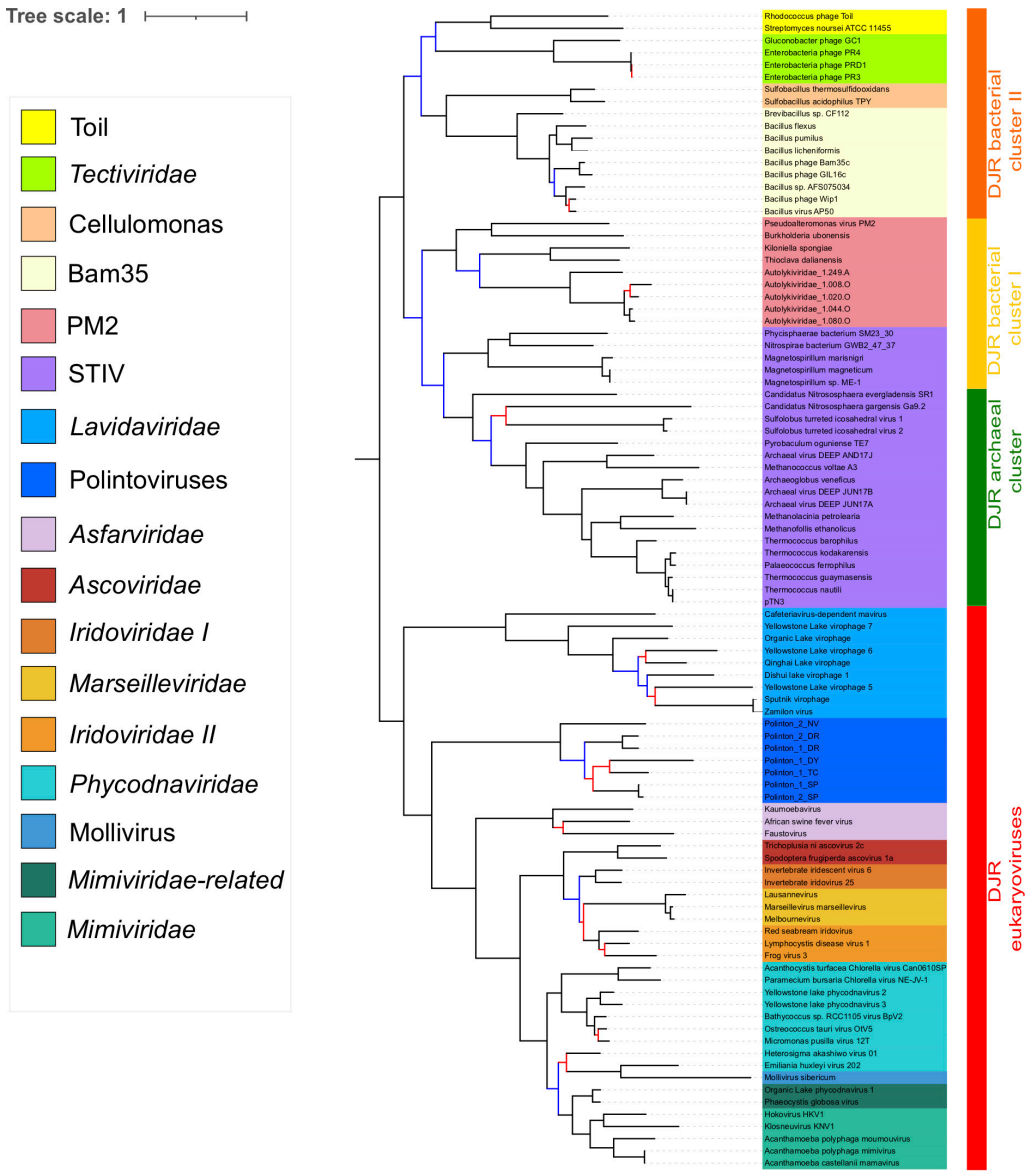


FIGURE 4 | Phylogenetic tree of the concatenated major capsid protein and packaging ATPase genes, excluding the *Poxviridae*. The alignment has 95 sequences with 434 positions. Phylogenetic tree was rooted between the prokaryotic and eukaryotic members. The scale-bar indicates the average number of substitutions per site. The best-fit model was LG + F + R5, which was chosen according to Bayesian Information Criterion (BIC). Branches in black indicate both classical bootstrap and transfer bootstrap expectation (TBE) support values using 1,000 replicates are above 70%. Branches in blue indicate only one of the two support values is above 70% whereas branches in red indicate both support values are below 70%.

sister clade to a clade grouping the archaeal *Tectiliviricetes* and bacterial *Tectiliviricetes* of the DJR bacterial cluster (**Figure 4** and **Supplementary Figure 7**).

Sphaerolipoviridae (*Helvetiavirae*) Branch With *Tectiliviricetes* (*Bamfordvirae*) in the pATPase Tree

As suggested by the pATPase amino-acid similarity network (**Supplementary Figure 3**) we could add the *Sphaerolipoviridae* pATPases sequences to our pATPase alignment. We thus obtained a pATPase tree in which *Sphaerolipoviridae* were grouped with *Tectiliviricetes*, as in our network analysis (**Figure 5** and **Supplementary Figure 8**). The relative position of the different groups of *Tectiliviricetes* in the pATPase tree remained somewhat similar to the same tree before the inclusion of *Sphaerolipoviridae* (**Figure 1**) except clear differences involving the STIV and PM2, probably due to a complex evolutionary pattern or a lack of resolution. *Sphaerolipoviridae* did not form a single monophyletic group basal to both eukaryoviruses and prokaryoviruses, as would have been expected in the SJR to DJR scenario, but two strongly supported monophyletic groups.

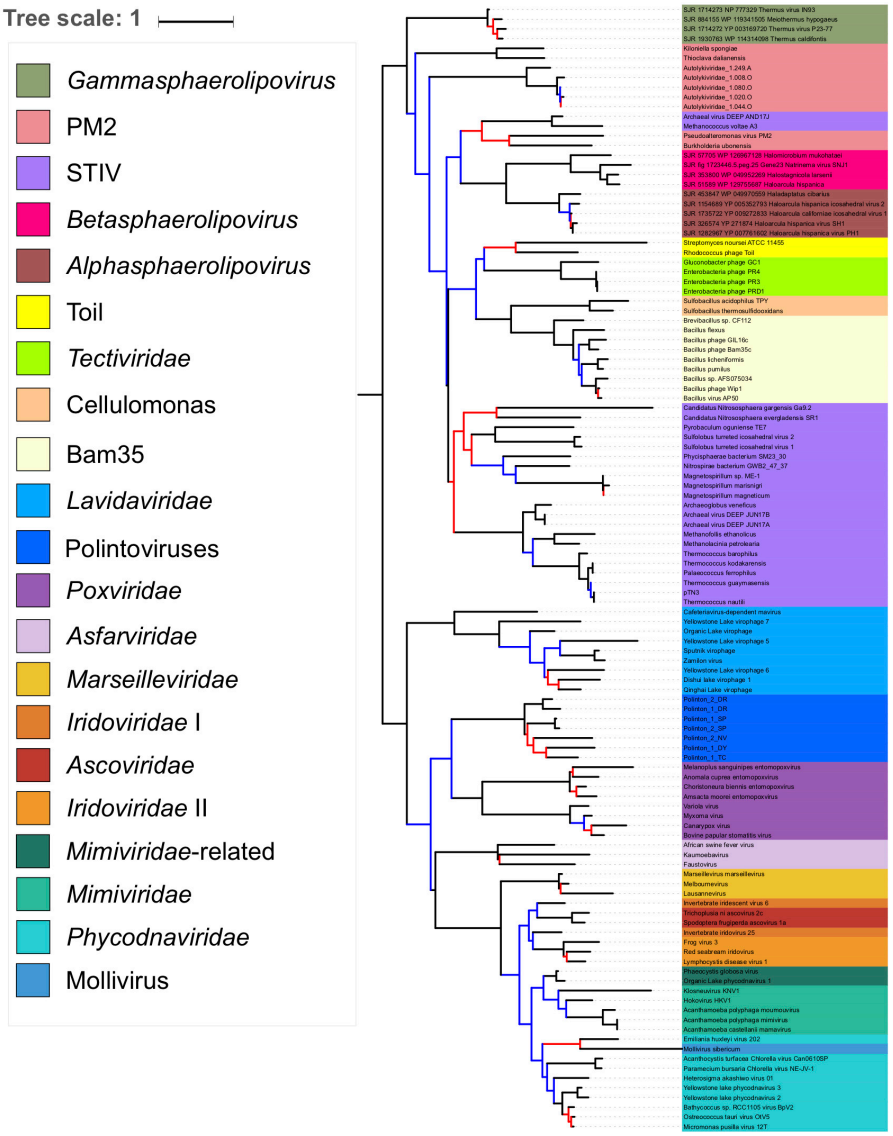


FIGURE 5 | Phylogenetic tree of the packaging ATPase gene of the *Varidnaviria* realm, excluding *Adenoviridae*. The alignment has 116 sequences with 151 positions. The root of the phylogenetic tree was between the prokaryotic and eukaryotic members. The scale-bar indicates the average number of substitutions per site. The best-fit model was LG + R6, which was chosen according to Bayesian Information Criterion (BIC). Branches in black indicate both classical bootstrap and transfer bootstrap expectation (TBE) support values using 1,000 replicates are above 70%. Branches in blue indicate only one of the two support values is above 70% whereas branches in red indicate both support values are below 70%.

Gamma Sphaerolipoviruses, which infect thermophilic bacteria, branched at the base of *Tectiliviricetes,* whereas Alpha and Beta Sphaerolipoviruses, which infect halophilic archaea branched deeper among *Tectiliviricetes*, forming a clade branching with two PM2 and two archaeal STIV sequences. This suggests two different origins for archaeal and bacterial SJR pATPases.

DISCUSSION

It has long been thought that it was not possible to build a valid sequence-based phylogeny of viruses infecting members of the three cellular domains. Here, we have obtained a rather well resolved and informative phylogeny for the realm *Varidnaviria*, based on the concatenation of their MCPs and pATPases. A similar strategy has been recently adopted to produce a global evolutionary history of the realm *Riboviria* (RNA viruses) based on the phylogeny of a single protein, their RNA-dependent RNA polymerases (Wolf et al., 2018). In the case of *Varidnaviria*, proteins involved in DNA replication cannot be used because different groups use non-homologous DNA replication proteins. However, a tree based on MCPs and pATPases might better correspond to what we expect for a "viral tree" if the virion and its mode of formation are considered to be the hallmark of

the virus (Bamford, 2003; Krupovic and Bamford, 2010; Forterre et al., 2014; Krupovic et al., 2020). We looked for the possibility to use additional proteins in virion morphogenesis, such as the minor capsid protein (mCP) and cysteine protease (PRO) to increase the robustness of our tree. However, mCPs and PROs are not well conserved in *Varidnaviria*. For instance, PROs have not been identified among *Tectiliviricetes*. The relatively small gene size of mCP and PRO also limits their usefulness in inferring phylogeny.

We could not include *Adenoviridae* and *Sphaerolipoviridae* in our concatenation because they did not encode the canonical MCP and/or pATPase. The MCP of *Adenoviridae* is too divergent from those of other *Varidnaviria*, whereas their pATPases belongs to another superfamily of P-loop ATPase (Burroughs et al., 2007). *Adenoviridae*, which have rather small genomes, have been tentatively included by the ICTV in the phylum *Preplasmiviricota* and the class *Tectiliviricetes* (Koonin et al., 2020) (**Table 1**). However, *Adenoviridae* exhibit more connection with *Nucleocytoviricota* than with *Tectiliviricetes* in a sequence similarity network (Sinclair et al., 2017) and branch far from *Tectiliviricetes* in a recent MCP structural tree (Ravantti et al., 2020), suggesting that classification of *Adenoviridae* within *Tectiliviricetes* could be premature.

In agreement with the classification of *Sphaerolipoviridae* as a distinct kingdom, *Helvetiavirae*, their MCP cannot be confidently aligned with those of *Bamfordvirae*. However, surprisingly, their pATPases could be aligned with those of *Bamfordvirae* and branched with *Tectiliviricetes* in the pATPase tree. Moreover, the pATPases of Alpha and Beta Sphaerolipoviruses and those of Gamma Sphaerolipoviruses branched at different positions (**Figure 5**). It was proposed that the single MCP with two jelly roll folds of *Bamfordvirae* originated by gene fusion of the two MCP of *Sphaerolipoviridae* (Krupovič and Bamford, 2008; Krupovic et al., 2020). However, our results suggest an alternative hypothesis, i.e., that the *Sphaerolipoviridae* SJR MCP originated twice from two distinct Tectivirus-like viruses by the deletion of one of the two jelly roll folds, followed by gene duplication. In that case, the MCP sequences of *Sphaerolipoviridae* might have rapidly diverged from those of *Bamfordvirae* following their structural rearrangements. An argument favoring the scenario from DJR to SJR might be the narrow distribution of *Sphaerolipoviridae*. In Archaea, *Sphaerolipoviridae* are only known infecting Haloarchaea, whereas in Bacteria, they only infect *Thermus* species, suggesting a "recent" emergence of these viruses. In contrast, the SJR to DJR scenario implies that *Sphaerolipoviridae* are very ancient, possibly predating LUCA (Krupovic et al., 2020). In that case, one would have expected a large distribution of these viruses in the three domains. Future exploration of the *Varidnaviria* diversity will possibly help to determine the correct scenario. In the meantime, it seems premature to consider the single family *Sphaerolipoviridae* as the prototype for a new kingdom.

Since we could not include *Sphaerolipoviridae* (*Helvetiavirae*) in our concatenated tree, this tree is formally a tree of *Bamfordvirae*. Noticeably, we recovered the monophyly of most families/groups previously defined on different criteria (Yutin et al., 2018). In particular, the internal phylogeny of the *Nucleocytoviricota*, with the monophyly of the order *Pimascovirales* (formerly the MAPI cluster) is very similar to the one that we previously obtained with eight-core genes of *Nucleocytoviricota*, the only difference being the position of *Asfarviridae* (Guglielmini et al., 2019). *Poxviridae* and *Asfarviridae* did not form a monophyletic group in our concatenated MCP/pATPase trees (**Figures 3**, **4**), in contradiction with findings of other studies, which recovered the clade with *Poxviridae* and *Asfarviridae* (Fischer et al., 2010; Hingamp et al., 2013), and their ICTV classification into the proposed class *Pokkesviricetes*. The *Asfarviridae* branch between Polintoviruses and *Nucleocytoviricota* in our *Bamfordvirae* tree without *Poxviridae* (**Figure 4**), whereas they emerged within Megaviricetes in our previous concatenation of the MCP and pATPase, which was limited to *Nucleocytoviricota* (except *Poxviridae*) and Polintoviruses (Guglielmini et al., 2019). The grouping of *Asfarviridae* with Megaviricetes was also observed in the MCP structural tree of Ravantti et al. (2020). We did not recover this grouping here and this is possibly due to long-branch attraction of *Asfarviridae* by the out-group sequences.

Our analysis supports the grouping of archaeal and bacterial *Bamfordvirae* in the same rank (*Tectiliviricetes*). There was a robust cluster (DJR cluster II) including the PRD1 (*Tectiviridae*), Bam35, Cellulomonas and Toil groups, which could correspond to the proposed ICTV order *Kalamavirales*. We also obtained the monophyly of archaeal STIV, which could correspond to the proposed ICTV order *Belfryvirales* and a robust cluster grouping members of the PM2 group and *Autolykiviridae*, as already suggested by Koonin and colleagues (Yutin et al., 2018). PM2 and relatives have been classified by the ICTV in the order *Vinavirales*. The position of bacterial STIV remains uncertain, they are sister group to the PM2 group in our concatenated tree with *Poxviridae* (**Figure 3**), suggesting classifying them in the order *Vinavirales* or a new order; in contrast, they form a monophyletic group with archaeal STIV in the concatenated tree without *Poxviridae* (**Figure 4**), suggesting to classify them in the order *Belfryvirales*.

Lavidaviridae (virophages) and Polintoviruses have been grouped with *Tectiliviricetes* in the same phylum, *Preplasmiviricota* based on a gene network analysis that has defined a Polinton-like module also including cytoplasmic and mitochondrial plasmids (Iranzo et al., 2016b) (hence the name *Preplasmiviricota*, *meaning* precursor of certain plasmids). This phylum was not recovered in our single-protein phylogenies, since either *Lavidaviridae* or Polintoviruses branch with *Poxviridae* that belong to *Nucleocytoviricota.* Although these branching are probably due to long branch attraction between *Poxviridae* and these two groups, it could also reflect a closer relationship of these two groups for *Nucleocytoviricota* than for *Tectiliviricetes,* as previously observed in a sequence similarity network (Sinclair et al., 2017). However, the grouping of *Lavidaviridae* or Polintoviruses with *Tectiliviricetes* is neither specifically supported nor refuted in our concatenated phylogeny since *Lavidaviridae* and Polintoviruses branched between *Tectiliviricetes* and *Nucleocytoviricota*.

We have previously shown that *Nucleocytoviricota* have already diverged before the Last Eukaryotic Common Ancestor (LECA) (Guglielmini et al., 2019). The present study indicates

that *Nucleocytoviricota*, *Lavidaviridae*, and Polintoviruses should have diverged even earlier and co-evolved for a long time with proto-eukaryotes. The early divergence of *Lavidaviridae* and *Nucleocytoviricota* is intriguing since all known modern *Lavidaviridae* (virophages) are parasites of *Imitervirales*. It suggests that the ancestors of modern *Lavidaviridae* used to infect proto-eukaryotes instead of giant viruses and were able to infect viruses present in their hosts subsequently. Integrated genomes of *Lavidaviridae* are abundant in some eukaryotes and used as tools to fight invading *Imitervirales* (Fischer and Hackl, 2016; Berjón-Otero et al., 2019). One can wonder if some *Lavidaviridae* are still able to infect eukaryotes in the absence of *Imitervirales* infection.

In **Figure 2**, we illustrate several of the possible scenarios for the evolution of *Bamfordvirae* (except *Adenoviridae*) and their implications for viral taxonomy, assuming that the structural module represents the vertical evolution of viruses. Koonin et al. (2006) proposed that *Tectiliviricetes* were already diversified at the time of LUCA and *Bamfordvirae* infecting eukaryotes evolved from a tectivirus that infected the bacterium at the origin of mitochondria (**Figure 2A**). They suggested that Polintoviruses originated first and became the ancestor of *Lavidaviridae* and *Nucleocytoviricota* (Krupovic and Koonin, 2015). Our MCP/pATPase concatenated tree does not support this specific version of their scenario since eukaryoviruses infecting *Bamfordvirae* did not branch within *Tectiviridae*. However, it is compatible with a rather similar scenario in which *Bamfordvirae* infecting eukaryotes evolved from an archaeal or a bacterial virus belonging to an extinct group of *Tectiliviricetes*. Notably, in such a scenario, the tree should be rooted within *Tectiliviricetes* and both *Tectiliviricetes* and *Preplasmiviricota* are paraphyletic (**Figure 2A**).

Our phylogenetic analysis produces a "viral tree of life" strikingly different from the cellular tree based on universal proteins in which either Archaea and eukaryotes are sister group or eukaryotes emerged within Archaea (**Figure 2B**) (Spang et al., 2015, 2018; Da Cunha et al., 2017, 2018), since *Bamfordvirae* infecting archaea and bacteria are grouped and separated from those infecting eukaryotes. In the scenario proposed by Koonin and colleagues, this contradiction is explained by the fact that modern *Tectiliviricetes* infecting archaea and bacteria have remained nearly identical to their ancestors 3–4 billion years ago, whereas the modern descendants of the tectivirus at the origin of *Bamfordvirae* infecting eukaryotes rapidly evolved into an immense variety of viral group, giving rise to *Lavidaviridae*, Polintoviruses and *Nucleocytoviricota* (**Figure 2A**).

The *Bamfordvirae* "viral tree of life" can be also explained in the framework of the classical Woese's tree of life. For instance, one can imagine that the ancestral *Bamfordvirae* common to archaea and eukaryotes were lost in the proto-archaeal lineage and replaced by *Bamfordvirae* of the bacterial type via mobilome transfer (**Figure 2B**). This scenario cannot be excluded since such mobilome transfer has been previously proposed in the case of conjugative plasmids (Guglielmini et al., 2013). The transfer of several components of the bacterial mobilome to proto-archaea could explain why the mobilomes of archaea and bacteria are very similar to each other while being very different from the eukaryotic mobilome (Forterre, 2013). In such a scenario, the concatenated tree could be rooted between *Tectiliviricetes* and *Bamfordvirae* infecting eukaryotes. Noticeably, *Nucleocytoviricota* and a clade corresponding to *Preplasmiviricota* except *Tectiliviricetes* (P*) form three monophyletic clades in that scenario (**Figure 2B**).

In the two scenarios previously discussed, we assumed that *Varidnaviria* were already present at the time of LUCA (**Figures 2A,B**). Another possibility is that *Varidnaviria* originated more recently (**Figures 2C,D**). In that case, considering the greater diversity of *Bamfordvirae* in Bacteria than in Archaea, it is tempting to imagine that these viruses originated in the bacterial lineage, suggesting a root within bacterial *Tectiliviricetes* (**Figure 2C**). If *Tectiliviricetes* were already present in the LACA as suggested by our analysis, this scenario again implies that both archaeal *Tectiliviricetes* and *Bamfordvirae* infecting eukaryotes originated from bacterial ones (**Figure 2C**). As in the case of the scenario of **Figure 2A**, *Tectiliviricetes* and *Preplasmiviricota* are both paraphyletic. Finally, an alternative version of a post-LUCA scenario is that *Bamfordvirae* originated in proto-eukaryotes and that some of them (related to Polintovirus/*Lavidaviridae*) were later on transferred to Bacteria, and finally from Bacteria to Archaea (**Figure 2D**). In that case, the MCP/ATPase tree could be rooted either between *Nucleocytoviricota* and *Preplasmiviricota* (including *Tectiliviricetes*) or within *Preplasmiviricota*, *Tectiliviricetes* forming a monophyletic group included in *Preplasmiviricota*. Although in contradiction with the current view suggesting that eukaryoviruses always originated from viruses infecting prokaryotes (Koonin et al., 2015), this hypothesis could explain why *Bamfordvirae* are so diverse and abundant in eukaryotes.

Interestingly, *Nucleocytoviricota* and *Preplasmiviricota* have strikingly different DNA replication proteins. In particular, *Nucleocytoviricota* share several of their major DNA replication proteins with head and tailed bacteriophages (*Caudovirales*) of the realm *Duplodnaviria* (Iranzo et al., 2016a). If *Nucleocytoviricota* originated from *Preplasmiviricota*, as suggested by Koonin and colleagues (**Figures 2A,C**), one should imagine that the DNA replication proteins encoded by the preplasmivirus at the origin of *Nucleocytoviricota* were replaced by the DNA replication proteins of *Caudovirales* infecting the same proto-eukaryotic hosts (black arrows on **Figure 2**). Such a replacement is also required in the scenario of **Figures 2B,D** (lower right panel). However, if the root of the *Bamfordvirae* tree is located between *Nucleocytoviricota* and *Preplasmiviricota* (**Figure 2D**, lower left panel), one can simply imagine that the divergence between these two phyla coincided with the association of an ancestral common virion morphogenesis module of the DJR type with two different types of replication modules, the replication modules that became associated with *Nucleocytoviricota* being a relative of the replication modules of some *Caudovirales*.

For some authors, the determination of viral phylogeny based on protein sequences comparison is a futile or at least risky exercise because they originated from different cell lines and that the core proteins characteristic of a modern viral lineage

could be an artifact due to the random losses of proteins initially present their cellular ancestors (Claverie, 2020; Nasir et al., 2020). This is probably not the case for the core proteins of *Nucleocytoviricota* since we have shown that the concatenation and single phylogenies of their eight-core genes were mostly congruent (Guglielmini et al., 2019). In any case, our result indicates that it is possible to trace the origin of the *Varidnaviria* virion morphogenesis module to a common ancestor that was already a virus, i.e., an organism that used the production of virions as the mode of dissemination of its genome (Raoult and Forterre, 2008), the origin and nature of this ancestral varidnavirus remaining enigmatic.

CONCLUSION

The presence of *Varidnaviria* in the three cellular domains raises challenging questions about their origin and evolution. Here, we have shown that phylogenies based on the concatenation of their MCP and pATPase can help to validate and/or question the viral classification and nomenclature of *Varidnaviria* recently proposed by the ICTV and can be used as a backbone to discuss current hypotheses about their evolution and propose new ones. In particular, we confirm the monophyly of *Tectiliviricetes* and *Nucleocytoviricota* and we identified a robust clade of *Tectiliviricetes* corresponding to the DJR cluster II. The tree presented here is not yet stable, as indicated by the fact that adding or removing some lineages impacts the relationships between some major clades. It will thus certainly be improved in the future with the discovery of new viral groups and the discovery of new members of the existing groups. Future identification and isolation of new viral families of *Varidnaviria*, especially in cellular lineages that have been poorly investigated until now, will thus be essential to possibly choose between the various scenarios for the history of this fascinating realm.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/ **Supplementary Material**.

AUTHOR CONTRIBUTIONS

AW and PF designed the study. AW, JG, and VD performed the bioinformatics experiments. AW, JG, MG, VD, and PF analyzed and interpreted the results. AW, MG, VD, and PF wrote the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by the European Research Council (ERC) grant from the European Union's Seventh Framework Program (FP/2007–2013)/Project EVOMOBIL-ERC Grant Agreement no. 340440.

ACKNOWLEDGMENTS

This work used the computational and storage services (TARS cluster) provided by the IT Department at Institut Pasteur, Paris.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2021.704052/full#supplementary-material

REFERENCES

Abrescia, N. G. A., Bamford, D. H., Grimes, J. M., and Stuart, D. I. (2012). Structure unifies the viral universe. *Annu. Rev. Biochem.* 81, 795–822. doi: 10.1146/annurev-biochem-060910-095130

Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W., et al. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 25, 3389–3402. doi: 10.1093/nar/25.17.3389

Baker, M. L., Jiang, W., Rixon, F. J., and Chiu, W. (2005). Common ancestry of herpesviruses and tailed DNA bacteriophages. *J. Virol.* 79, 14967–14970. doi: 10.1128/jvi.79.23.14967-14970.2005

Bamford, D. H. (2003). Do viruses form lineages across different domains of life? *Res. Microbiol.* 154, 231–236. doi: 10.1016/S0923-2508(03)00065-62

Bamford, D. H., Grimes, J. M., and Stuart, D. I. (2005). What does structure tell us about virus evolution? *Curr. Opin. Struct. Biol.* 15, 655–663. doi: 10.1016/j.sbi.2005.10.012

Berjón-Otero, M., Koslová, A., and Fischer, M. G. (2019). The dual lifestyle of genome-integrating virophages in protists. *Ann. N. Y. Acad. Sci.* 1447, 97–109. doi: 10.1111/nyas.14118

Burroughs, A., Iyer, L., and Aravind, L. (2007). Comparative genomics and evolutionary trajectories of viral ATP dependent DNA-packaging systems. *Genome Dyn.* 3, 48–65. doi: 10.1159/000107603

Claverie, J. M. (2020). Fundamental difficulties prevent the reconstruction of the deep phylogeny of viruses. *Viruses* 12:1130. doi: 10.3390/v12101130

Da Cunha, V., Gaia, M., Gadelle, D., Nasir, A., and Forterre, P. (2017). Lokiarchaea are close relatives of Euryarchaeota, not bridging the gap between prokaryotes and eukaryotes. *PLoS Genet.* 13:e1006810. doi: 10.1371/journal.pgen.1006810

Da Cunha, V., Gaia, M., Nasir, A., and Forterre, P. (2018). Asgard archaea do not close the debate about the universal tree of life topology. *PLoS Genet.* 14:e1007215. doi: 10.1371/journal.pgen.1007215

Demina, T. A., Pietilä, M. K., Svirskaitë, J., Ravantti, J. J., Atanasova, N. S., Bamford, D. H., et al. (2017). HCIV-1 and other tailless icosahedral internal membrane-containing viruses of the family Sphaerolipoviridae. *Viruses* 9:32. doi: 10.3390/v9020032

Fischer, M. G., Allen, M. J., Wilson, W. H., and Suttle, C. A. (2010). Giant virus with a remarkable complement of genes infects marine zooplankton. *Proc. Natl. Acad. Sci. U S A.* 107, 19508–19513. doi: 10.1073/pnas.1007615107

Fischer, M. G., and Hackl, T. (2016). Host genome integration and giant virus-induced reactivation of the virophage mavirus. *Nature* 540, 288–291. doi: 10.1038/nature20593

Forterre, P. (2013). The common ancestor of archaea and eukarya was not an archaeon. *Archaea* 2013:372396. doi: 10.1155/2013/372396

Forterre, P., Krupovic, M., and Prangishvili, D. (2014). Cellular domains and viral lineages. *Trends Microbiol.* 22, 554–558. doi: 10.1016/j.tim.2014.07.004

Gaudin, M., Krupovic, M., Marguet, E., Gauliard, E., Cvirkaite-Krupovic, V., Le Cam, E., et al. (2014). Extracellular membrane vesicles harbouring viral genomes. *Environ. Microbiol.* 16, 1167–1175. doi: 10.1111/1462-2920.12235

Gil-Carton, D., Jaakkola, S. T., Charro, D., Peralta, B., Castaño-Díez, D., Oksanen, H. M., et al. (2015). Insight into the assembly of viruses with vertical single β-barrel major capsid proteins. *Structure* 23, 1866–1877. doi: 10.1016/j.str.2015.07.015

Guglielmini, J., De La Cruz, F., and Rocha, E. P. C. (2013). Evolution of conjugation and type IV secretion systems. *Mol. Biol. Evol.* 30, 315–331. doi: 10.1093/molbev/mss221

Guglielmini, J., Woo, A. C., Krupovic, M., Forterre, P., and Gaia, M. (2019). Diversification of giant and large eukaryotic dsDNA viruses predated the origin of modern eukaryotes. *Proc. Natl. Acad. Sci. U S A.* 116, 19585–19592. doi: 10.1073/pnas.1912006116

Hennig, W. (1965). Phylogenetic systematics. *Annu. Rev. Entomol.* 10, 97–116. doi: 10.1146/annurev.en.10.010165.000525

Hingamp, P., Grimsley, N., Acinas, S. G., Clerissi, C., Subirana, L., Poulain, J., et al. (2013). Exploring nucleo-cytoplasmic large DNA viruses in Tara Oceans microbial metagenomes. *ISME J.* 7, 1678–1695. doi: 10.1038/ismej.2013.59

Iranzo, J., Koonin, E. V., Prangishvili, D., and Krupovic, M. (2016a). Bipartite network analysis of the archaeal virosphere: evolutionary connections between viruses and capsidless mobile elements. *J. Virol.* 90, 11043–11055. doi: 10.1128/jvi.01622-1616

Iranzo, J., Krupovic, M., and Koonin, E. V. (2016b). The double-Stranded DNA virosphere as a modular hierarchical network of gene sharing. *mBio* 7:e00978-16. doi: 10.1128/mbio.00978-916

Jurka, J., Kapitonov, V. V., Pavlicek, A., Klonowski, P., Kohany, O., and Walichiewicz, J. (2005). Repbase update, a database of eukaryotic repetitive elements. *Cytogenet. Genome Res.* 110, 462–467. doi: 10.1159/000084979

Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K. F., Von Haeseler, A., and Jermiin, L. S. (2017). ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat. Methods* 14, 587–589. doi: 10.1038/nmeth.4285

Katoh, K., and Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* 30, 772–780. doi: 10.1093/molbev/mst010

Kauffman, K. M., Hussain, F. A., Yang, J., Arevalo, P., Brown, J. M., Chang, W. K., et al. (2018). A major lineage of non-tailed dsDNA viruses as unrecognized killers of marine bacteria. *Nature* 554, 118–112. doi: 10.1038/nature25474

Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N., and Sternberg, M. J. E. (2015). The Phyre2 web portal for protein modeling, prediction and analysis. *Nat. Protoc.* 10, 845–858. doi: 10.1038/nprot.2015.053

Koonin, E., Dolja, V., Krupovic, M., Varsani, A., Wolf, Y., Yutin, N., et al. (2019). *Create a Megataxonomic Framework, Filling all Principal Taxonomic Ranks, for DNA Viruses Encoding Vertical Jelly Roll-type Major Capsid Proteins*. London: International Committee on Taxonomy of Viruses, doi: 10.13140/RG.2.2.14886.47684

Koonin, E. V., Dolja, V. V., and Krupovic, M. (2015). Origins and evolution of viruses of eukaryotes: the ultimate modularity. *Virology* 479–480, 2–25. doi: 10.1016/j.virol.2015.02.039

Koonin, E. V., Dolja, V. V., Krupovic, M., Varsani, A., Wolf, Y. I., Yutin, N., et al. (2020). Global organization and proposed megataxonomy of the virus world. *Microbiol. Mol. Biol. Rev.* 84:2. doi: 10.1128/mmbr.00061-19

Koonin, E. V., Senkevich, T. G., and Dolja, V. V. (2006). The ancient virus world and evolution of cells. *Biol. Direct.* 1:29. doi: 10.1186/1745-6150-1-29

Koonin, E. V., and Yutin, N. (2019). Evolution of the large nucleocytoplasmic DNA viruses of eukaryotes and convergent origins of viral gigantism. *Adv. Virus Res.* 103, 167–202. doi: 10.1016/bs.aivir.2018.09.002

Krupovič, M., and Bamford, D. H. (2008). Virus evolution: how far does the double β-barrel viral lineage extend? *Nat. Rev. Microbiol.* 6, 941–948. doi: 10.1038/nrmicro2033

Krupovic, M., and Bamford, D. H. (2010). Order to the viral universe. *J. Virol.* 84, 12476–12479. doi: 10.1128/jvi.01489-1410

Krupovic, M., Bamford, D. H., and Koonin, E. V. (2014). Conservation of major and minor jelly-roll capsid proteins in Polinton (Maverick) transposons suggests that they are bona fide viruses. *Biol. Direct.* 9:6. doi: 10.1186/1745-6150-9-6

Krupovic, M., Dolja, V. V., and Koonin, E. V. (2020). The LUCA and its complex virome. *Nat. Rev. Microbiol.* 18, 661–670. doi: 10.1038/s41579-020-0408-x

Krupovic, M., and Koonin, E. V. (2015). Polintons: a hotbed of eukaryotic virus, transposon and plasmid evolution. *Nat. Rev. Microbiol.* 13, 105–115. doi: 10.1038/nrmicro3389

Krupovic, M., Makarova, K. S., Wolf, Y. I., Medvedeva, S., Prangishvili, D., Forterre, P., et al. (2019). Integrated mobile genetic elements in thaumarchaeota. *Environ. Microbiol.* 21, 2056–2078. doi: 10.1111/1462-2920.14564

Laanto, E., Mäntynen, S., De Colibus, L., Marjakangas, J., Gillum, A., Stuart, D. I., et al. (2017). Virus found in a boreal lake links ssDNA and dsDNA viruses. *Proc. Natl. Acad. Sci. U S A.* 114, 8378–8383. doi: 10.1073/pnas.1703834114

Lemoine, F., Domelevo Entfellner, J. B., Wilkinson, E., Correia, D., Dávila Felipe, M., De Oliveira, T., et al. (2018). Renewing Felsenstein's phylogenetic bootstrap in the era of big data. *Nature* 556, 452–456. doi: 10.1038/s41586-018-0043-40

Letunic, I., and Bork, P. (2007). Interactive Tree Of Life (iTOL): an online tool for phylogenetic tree display and annotation. *Bioinformatics* 23, 127–128. doi: 10.1093/bioinformatics/btl529

Miele, V., Penel, S., and Duret, L. (2011). Ultra-fast sequence clustering from similarity networks with SiLiX. *BMC Bioinformatics* 12:116. doi: 10.1186/1471-2105-12-116

Nasir, A., Romero-Severson, E., and Claverie, J. M. (2020). Investigating the concept and origin of viruses. *Trends Microbiol.* 28, 959–967. doi: 10.1016/j.tim.2020.08.003

Nguyen, L. T., Schmidt, H. A., Von Haeseler, A., and Minh, B. Q. (2015). IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* 32, 268–274. doi: 10.1093/molbev/msu300

Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., et al. (2004). UCSF Chimera - a visualization system for exploratory research and analysis. *J. Comput. Chem.* 25, 1605–1612. doi: 10.1002/jcc.20084

Raoult, D., and Forterre, P. (2008). Redefining viruses: lessons from Mimivirus. *Nat. Rev. Microbiol.* 6, 315–319. doi: 10.1038/nrmicro1858

Ravantti, J., Bamford, D., and Stuart, D. I. (2013). Automatic comparison and classification of protein structures. *J. Struct. Biol.* 183, 47–56. doi: 10.1016/j.jsb.2013.05.007

Ravantti, J. J., Martinez-Castillo, A., and Abrescia, N. G. A. (2020). Superimposition of viral protein structures: a means to decipher the phylogenies of viruses. *Viruses* 12:1146. doi: 10.3390/v12101146

San Martín, C., and van Raaij, M. J. (2018). The so far farthest reaches of the double jelly roll capsid protein fold. *Virol. J.* 15:181. doi: 10.1186/s12985-018-1097-1091

Sinclair, R., Ravantti, J., and Bamford, D. H. (2017). Nucleic and amino acid sequences classification. *J. Virol.* 91:e02275-16. doi: 10.1128/JVI.02275-2216

Spang, A., Eme, L., Saw, J. H., Caceres, E. F., Zaremba-Niedzwiedzka, K., Lombard, J., et al. (2018). Asgard archaea are the closest prokaryotic relatives of eukaryotes. *PLoS Genet.* 14:e1007080. doi: 10.1371/journal.pgen.1007080

Spang, A., Saw, J. H., Jørgensen, S. L., Zaremba-Niedzwiedzka, K., Martijn, J., Lind, A. E., et al. (2015). Complex archaea that bridge the gap between prokaryotes and eukaryotes. *Nature* 521, 173–179. doi: 10.1038/nature14447

Wang, H. C., Minh, B. Q., Susko, E., and Roger, A. J. (2018). Modeling site heterogeneity with posterior mean site frequency profiles accelerates accurate phylogenomic estimation. *Syst. Biol.* 67, 216–235. doi: 10.1093/sysbio/syx068

Woese, C. R., Kandler, O., and Wheelis, M. L. (1990). Towards a natural system of organisms: proposal for the domains archaea. bacteria, and eucarya. *Proc. Natl. Acad. Sci. U S A.* 87, 4576–4579. doi: 10.1073/pnas.87.12.4576

Wolf, Y. I., Kazlauskas, D., Iranzo, J., Lucía-Sanz, A., Kuhn, J. H., Krupovic, M., et al. (2018). Origins and evolution of the global RNA virome. *mBio* 9:e02329-18. doi: 10.1128/mbio.02329-2318

Yutin, N., Bäckström, D., Ettema, T. J. G., Krupovic, M., and Koonin, E. V. (2018). Vast diversity of prokaryotic virus genomes encoding double jelly-roll major capsid proteins uncovered by genomic and metagenomic sequence analysis. *Virol. J.* 15:67. doi: 10.1186/s12985-018-0974-y

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Microbiology

ORIGINAL RESEARCH
published: 10 August 2021
doi: 10.3389/fmicb.2021.715608

Clandestinovirus: A Giant Virus With Chromatin Proteins and a Potential to Manipulate the Cell Cycle of Its Host *Vermamoeba vermiformis*

Clara Rolland[1,2], Julien Andreani[1], Dehia Sahmi-Bounsiar[1,2], Mart Krupovic[3], Bernard La Scola[1,2] and Anthony Levasseur[1,2,4]**

[1] Aix-Marseille Université (AMU), UMR MEPHI (Microbes, Evolution, Phylogeny and Infections), IRD, APHM, Faculté de Médecine, Marseille, France, [2] IHU Méditerranée Infection, Marseille, France, [3] Archaeal Virology Unit, Institut Pasteur, Paris, France, [4] Institut Universitaire de France, Paris, France

OPEN ACCESS

Edited by:
Erna Geessien Kroon, Federal University of Minas Gerais, Brazil

Reviewed by:
Juliana Cortines, Federal University of Rio de Janeiro, Brazil
Mohammad Moniruzzaman, Virginia Tech, United States

***Correspondence:**
Bernard La Scola
bernard.la-scola@univ-amu.fr
Anthony Levasseur
anthony.levasseur@univ-amu.fr

Specialty section:
This article was submitted to Virology, a section of the journal Frontiers in Microbiology

Received: *27 May 2021*
Accepted: *14 July 2021*
Published: *10 August 2021*

Citation:
Rolland C, Andreani J, Sahmi-Bounsiar D, Krupovic M, La Scola B and Levasseur A (2021) Clandestinovirus: A Giant Virus With Chromatin Proteins and a Potential to Manipulate the Cell Cycle of Its Host Vermamoeba vermiformis. Front. Microbiol. 12:715608. doi: 10.3389/fmicb.2021.715608

For several decades, the vast world of DNA viruses has been expanding constantly. Various discoveries in this field have broadened our knowledge and revealed that DNA viruses encode many functional features, which were once thought to be exclusive to cellular life. Here, we report the isolation of a giant virus named "clandestinovirus," grown on the amoebal host *Vermamoeba vermiformis*. This virus was discovered in a mixed co-culture associated with another giant virus, Faustovirus ST1. Clandestinovirus possesses a linear dsDNA genome of 581,987 base pairs containing 617 genes. Phylogenetically, clandestinovirus is most closely related to Acanthamoeba castellanii medusavirus and was considered a member of the proposed *Medusaviridae* family. However, clandestinovirus genome is 65% larger than that of medusavirus, emphasizing the considerable genome size variation within this virus family. Functional annotation of the clandestinovirus genes suggests that the virus encodes four core histones. Furthermore, clandestinovirus appears to orchestrate the cell cycle and mitochondrial activities of the infected host by virtue of encoding a panel of protein kinases and phosphatases, and a suite of functionally diverse mitochondrial protein homologs, respectively. Collectively, these observations illuminate a strategy employed by clandestinovirus to optimize the intracellular environment for efficient virus propagation.

Keywords: giant virus, nucleocytoviricota, clandestinovirus, histones, mitochondria, amoeba

INTRODUCTION

Exploration of the giant viruses in amoeba began in 2003 with the discovery of *Acanthamoeba polyphaga mimivirus* (La Scola et al., 2003). Currently, the giant viruses are classified into the phylum *Nucleocytoviricota*, previously designated as Nucleo-cytoplasmic large DNA viruses (NCLDV) (Iyer et al., 2001; Walker et al., 2019; Koonin et al., 2020). Due to their large genome and virion sizes, and unprecedented gene composition, the giant viruses, in many ways, have redefined the whole concept of a "virus." Indeed, some giant viruses can possess thousands of genes (Colson et al., 2011; Philippe et al., 2013) with, notably, elements of the translation apparatus, such as tRNA,

aminoacyl tRNA synthetase (aaRS) or translation factors (Raoult et al., 2004; Schulz et al., 2017; Abrahão et al., 2018). Recent progress in metagenomics and sequence assembly made it possible to detect and reconstruct *in silico* complete genomes of giant viruses from diverse environments (Moniruzzaman et al., 2020a; Schulz et al., 2020b). This contribution enabled the discovery of unexpected metabolic pathways and functions, which were once thought not to be encoded by viruses. Indeed, channelrhodopsins and key enzymatic pathways, such as tricarboxylic cycle, traditionally attributed to cellular life, were discovered in viral genomes, revealing the potential roles played by those giant viruses in their ecosystems (Yutin and Koonin, 2012; Needham et al., 2019a,b). Moreover, giant viruses have been shown to be integrated in the genomes of green algae (Moniruzzaman et al., 2020b) and fungi (Gong et al., 2020). This phenomenon appears to be widespread and highlights the need to further characterize those viruses in order to fully understand their global contributions to the evolution and ecology of eukaryotes.

Some viruses, namely, those belonging to the *Marseilleviridae* family, encode histone doublets in their genomes (Boyer et al., 2009; Thomas et al., 2011). Recently, it has been demonstrated experimentally that histones encoded by two different marseilleviruses localize to the virus factories in the cell cytoplasm and form nucleosomes remarkably similar to those of eukaryotes (Liu Y. et al., 2021; Valencia-Sánchez et al., 2021). Acanthamoeba castellanii medusavirus (medusavirus), a giant virus discovered in a Japanese hot spring, and medusavirus stheno were also shown to possess the four core histones and a linker histone H1 (Yoshikawa et al., 2019; Yoshida et al., 2021). Notably, unlike most other giant viruses, including marseilleviruses, which replicate exclusively in the cytoplasm, replication of the medusavirus genome takes place in the host nucleus (Yoshikawa et al., 2019).

In 2019, we reported the isolation of a potential novel virus, which was mixed with faustovirus ST1 (Rolland et al., 2019). Using single-cell microaspiration applied on amoeba, we succeeded in separating the two viruses (Sahmi-Bounsiar et al., 2019). Here, we describe this new giant virus, named clandestinovirus, discovered in *Vermamoeba vermiformis*. Clandestinovirus shares genes and displays phylogenetic relationship to the Acanthamoeba castellanii medusavirus, suggesting that it is a distant new member of the proposed *Medusaviridae* family. Among the notable features of clandestinovirus, we highlight the presence of genes encoding histones and proteins involved in histone tail modification.

MATERIALS AND METHODS

Sample Collection, Virus Isolation, Production and Purity Control

The co-culture method associated with the single cell micro-aspiration was used to isolate Clandestinovirus in sample from wastewater in Saint-Pierre-de-Mézoargues, France. *Vermamoeba vermiformis* (strain CDC19) was used as cell support. After 48 h in Peptone Yeast extract Glucose medium (PYG) amoebas were washed, harvested at a concentration of 1×10^6 cells/ml and pelleted at 720 × g for 10 min. The cells were re-suspended in the starvation medium at a concentration of 1×10^6 cells/ml (Bou Khalil et al., 2016). Then, the viral sample was inoculated on the cell support at a multiplicity of infection (MOI) of 0.01 and incubated at 30°C until cytopathic effects (CPE) were induced. At this stage, Clandestinovirus was present in our culture but mixed with faustovirus ST1 (Louazani et al., 2017). To separate the two viruses, we used the single cell micro-aspiration method (Sahmi-Bounsiar et al., 2019). Prior to the micro-aspiration, an end-point dilution (10^{-1} to 10^{-11}) of the mixed viruses was made in Petri dishes containing 2 ml of amoebas at 1×10^6 cells/ml and 100 µl of the viral mix. At the appearance of the rounding of cells (CPE), the micro-aspiration was performed. Briefly, single infected amoebas from the prepared Petri dishes was picked and cloned into a new Petri dish containing a fresh monolayer of *V. vermiformis*, then incubated at 30°C. We monitored the emergence of cytopathic effect characterized by rounding cells daily with an inverted optical microscope and screened the positive culture by standard PCR using specific primers against faustovirus (targeting the *rpb2* gene) and against clandestinovirus (minor capsid protein encoding gene). When we succeeded in separating the two viruses, we proceeded to the viral strain production of clandestinovirus. In brief, 15 flasks of 150 cm^2 (Corning®, NY, United States), infected with the positive culture of clandestinovirus only, were pelleted using the Beckman coulter® Optima™ XPN-80 ultracentrifuge (Beckman Coulter, France) at 50,000 × *g* for 45 min. We finished the purification of the virus using a sucrose gradient of 25% in order to perform whole genome sequencing.

Replicative Cycle

We use the same protocol as with cedratvirus (Andreani et al., 2016) for the infectious cycle description and the electron microscopy. Briefly, five flasks each containing 30 ml of amoebas at a concentration of 5×10^5 cells/ml were infected with the pure viral suspension at a MOI of 10. After 40 min of incubation, amoebas were washed three times with PAS buffer to eliminate non-internalized viruses. This time point was designated as H0. New flasks were filled with 10 ml of the infected cultures and incubated at 30°C. A flask of non-infected amoebas was used as negative control. At 0, 2, 4, 6, 7, 8, 9, 10, 11, 12 13, 14, 16, 20, and 24 h post infection (hpi), each culture flask corresponding to the specific time point was centrifugated at 720 × *g* for 10 min, and the pellets were fixed for the transmission electron microscopy.

Genome Sequencing and Genome Assembly

Genomic DNA (gDNA) of clandestinovirus was extracted in two steps: a mechanical treatment was first performed by glass beads acid washed (G4649-500g Sigma) using a FastPrep™-24 5G Grinder (mpBio) at maximum speed (6.5) for 90 s. Then after 30 min lysozyme incubation at 37°C, DNA was extracted on the EZ1 biorobot (Qiagen) with EZ1 DNA tissues kit. The elution volume is 50 µL. gDNA of clandestinovirus

was quantified by a Qubit assay with the high sensitivity kit (Life technologies, Carlsbad, CA, United States) to 0.2 ng/μl. Genomic DNA was next sequenced on the MiSeq Technology (Illumina Inc., San Diego, CA, United States) with the paired end strategy and was barcoded in order to be mixed, respectively, with 24 other genomic projects prepared with the Nextera XT DNA sample prep kit (Illumina). To prepare the paired end library, dilution was performed to require 1 ng of each genome as input to prepare the paired end library. The tagmentation step fragmented and tagged the DNA. Then limited cycle PCR amplification (12 cycles) completed the tag adapters and introduced dual-index barcodes. After purification on AMPure XP beads (Beckman Coulter Inc., Fullerton, CA, United States), the libraries were then normalized on specific beads according to the Nextera XT protocol (Illumina). Normalized libraries were pooled into a single library for sequencing on the MiSeq. The pooled single strand library was loaded onto the reagent cartridge and then onto the instrument along with the flow cell. Automated cluster generation and paired end sequencing with dual index reads were performed in a single 39-h run in 2 × 250-bp. Total information of 4.97 Gb was obtained from a 536,000 cluster density per mm^2 with a cluster passing quality control filters of 94.32%. Within this run, the index representation for clandestinovirus was determined to index 5.78%. The 10,397,314 paired end reads were filtered according to the read qualities.

The genome was assembled using HybridSpades using the default parameters (Antipov et al., 2016).

Genome Analysis

Gene prediction was computed using GenemarkS (Besemer et al., 2001) and Prodigal v2.6.2 software (Hyatt et al., 2010). The predicted proteins that were smaller than 50 amino acids were deleted. Those between 50 and 99 amino acids were analyzed with Phyre2 software and eliminated of the dataset if they had an abnormal tri-dimensional folding (i.e., proteins with a confidence cut-off under 70% were discarded) (Kelley et al., 2015). A BlastP analysis was performed against the non-redundant (nr) protein database (5 July 2020) using an e-value cut-off of 1E-3. The annotation was performed using a combination of CD-search tool (Marchler-Bauer et al., 2017) and Interpro version 81.0 (Mitchell et al., 2019). In addition, sensitive profile-profile comparisons were performed using HHsearch against the pfam, CDD and COG databases (Steinegger et al., 2019). The core genome analysis was conducted using ProteinOrtho v6 with the parameters of 60% coverage and 20% identity (Lechner et al., 2011). The comparison was performed between clandestinovirus and other giant viruses classified by family. The virus groups included were Acanthamoeba castellanii medusavirus, *Mimiviridae*, klosneuviruses, *Marseilleviridae*, *Pithoviridae*, pandoraviruses, and molliviruses in one group, and *Asfarviridae*, faustoviruses, kaumoebavirus, and pacmanvirus A23 as another group (**Supplementary Table 1**).

Phylogenetic Analysis

All phylogenetic analyses were performed using the following procedure. A BlastP with the predicted protein limited to viruses in the NCBI non-redundant (nr) protein database, using an *e*-value cut-off of 1E-3, was computed to retrieve homologous proteins. We kept an average of 65 hits per query protein (ranging between 78 and 57) except for the A32 packaging ATPase and RPB5, where only 49 and 38 homologous proteins were retrieved. For the phylogeny of histones, BlastP was conducted against the non-redundant protein database retaining around 30 hits. Then Muscle program was used to align the sequences (Edgar, 2004). Phylogenetic trees were constructed with the FastTree software using the Jones-Taylor-Thornton (JTT) model for amino acid substitution and the maximum likelihood method with 1,000 bootstrap replications (Price et al., 2010). The trees were visualized using iTOL v6 (Letunic and Bork, 2021).

RESULTS

Viral Isolation

The sequencing of the faustovirus ST1 genome revealed the presence of a second genome in the sample. The classical method of giant virus isolation by end-point dilution was used to separate the two viruses in co-culture but without success. A new, recently developed technique, single cell micro-aspiration, was then used allowing the separation of faustovirus ST1 and the second virus, named clandestinovirus (Sahmi-Bounsiar et al., 2019; **Supplementary Figure 1**). After separation, the new virus was cultivated on *V. vermiformis*, the only known host of Faustovirus, but no lysis was observed. Instead, the infection of *V. vermiformis* by clandestinovirus led to rounding of the cells around 8 h post-infection (hpi.) and their detachment after 13 hpi.

Description of the Replication Cycle

Transmission electron microscopy was used to measure viral particles and describe the replicative cycle of clandestinovirus. The virus has an icosahedral capsid without fibrils. The virion size ranges between 175 and 202 nm with a mean size of 180 nm (SD ± 12 nm; $N = 71$). Briefly, the replicative cycle begin with the entry of the virus into amoeba by phagocytosis (**Figures 1A,B**). Once inside the cell, between 4 and 7 hpi., the virus migrates through the cytoplasm to the nucleus. At this stage, the clandestinovirus particles cling against the nuclear membrane before the entry of the virus into the nucleus (**Figures 1C,D**). The replication takes place in the nucleus, which is converted into a viral factory, from 7 to 12 hpi. (**Figures 1E–G**) as it could observed for Medusavirus in *Acanthamoeba* (Yoshikawa et al., 2019). We clearly see the filing of capsids in the **Figure 1G**. From 10 hpi., mature virus particles accumulate outside the viral factory, in the cytoplasm of the host *V. vermiformis* (**Figures 1H,I**). These areas of accumulation of mature particles were not observed during the Medusavirus infection where virions were more dispersed in the cytoplasm. The newly assembled virions were released by exocytosis around 16 hpi. In the course of infection, the cells became rounded, lost their adhesion and remained in this state even after 7 days of infection monitored by inverted microscope.

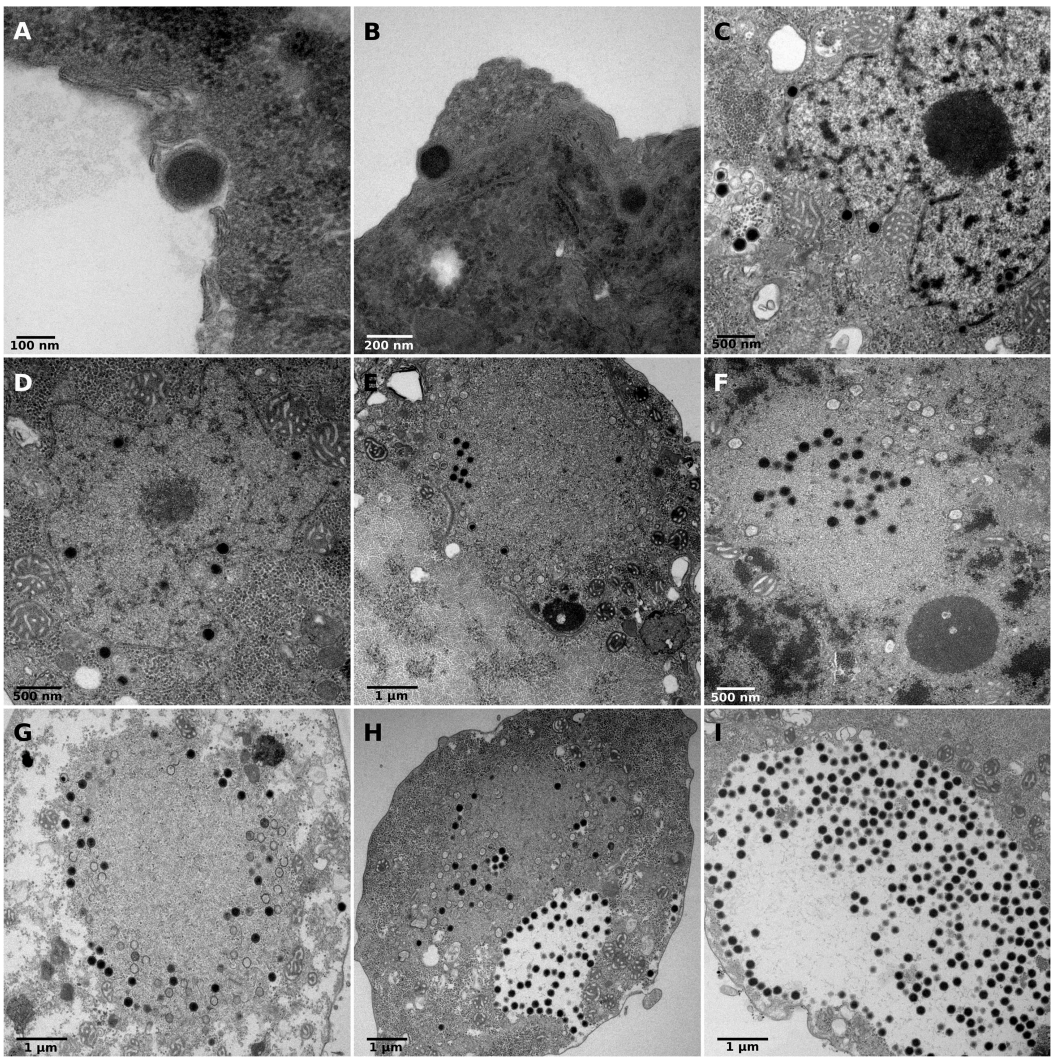


FIGURE 1 | Replicative cycle of clandestinovirus. Scale bars are indicated on each panels. **(A,B)** Entry of the virus into *Vermamoeba vermiformis* by phagocytosis at 4 hpi. **(C,D)** Migration in the cytoplasm to the nucleus and entry into the host nucleus at 7 and 8 hpi. **(E–G)** Shows the replication of clandestinovirus in the nucleus which became a viral factory at 10 hpi **(E)** and 12 hpi **(F,G)**. **(H,I)** Accumulation of mature viral particles at 11 and 12 hpi.

Genomic Characterization of Clandestinovirus

Clandestinovirus possesses a linear genome of 581,987 base pairs (bp) composed of two scaffolds of 562,042 and 19,445 bp with a G + C content estimated at 43.5 and 46%, respectively (**Table 1**). A BlastN search of the whole genome predicted 652 ORF but 35 were discarded due to an abnormal tri-dimensional folding (proteins with predictive 3D structure under a confidence cut-off of 70% were removed). Ultimately, 617 genes were retained with a distribution between the two scaffolds of 559 on scaffold 1 and 17 on scaffold 2, corresponding to a global coding proportion of 86.8% of the whole genome (504,963 bp). Two tRNAs were also retrieved, one was a serine tRNA (Ser-tRNA) and the other was a pseudo tRNA with 32% similarity to histidine tRNA (His-tRNA).

BlastP against the nr database yielded 214 proteins with at least one homolog (≈34.7% of all predicted proteins) and 403 unmatched proteins, classified as ORFans (≈65.3% of all predicted genes). Of the 214 proteins, 86 had the best hit with viruses (≈40%), 81 with eukaryotes (≈38%) and 47 with prokaryotes (≈22%; 44 best hits with bacteria and three with archaea) (**Figure 2A**). For the viral hits, 32% of

TABLE 1 | Genomic characteristics of clandestinovirus.

Main genomic characteristics	
Genome size (bp)	581,987
GC content (%)	43.5
Predicted proteins	617
ORFans (%)	≈ 65.3
Coding density	86.80%
tRNA	2

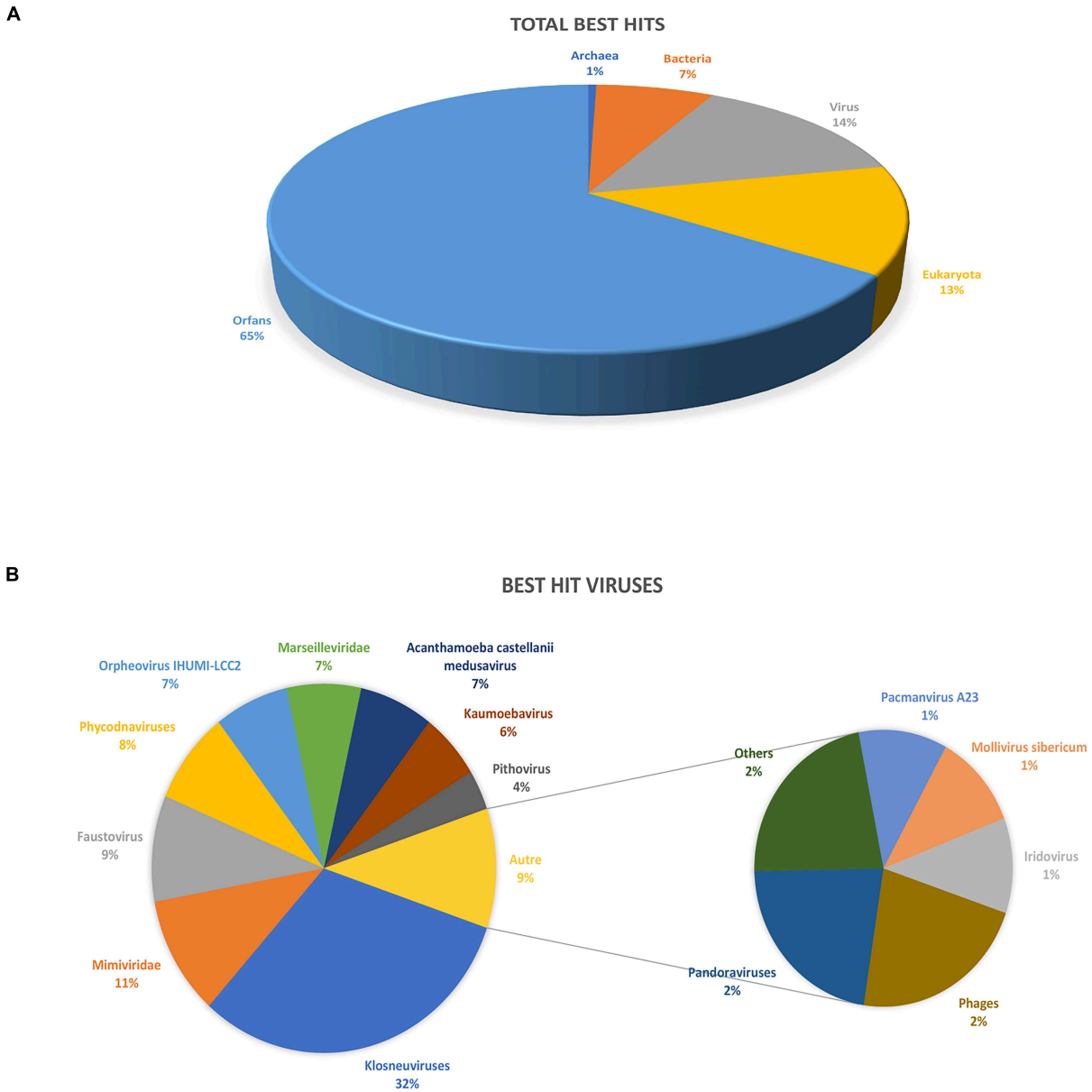


FIGURE 2 | Description of clandestinovirus best hits obtained by blastp against the NCBI non-redundant database. **(A)** Total number of best hits by blast. **(B)** Description of the viruses' best hits.

the best hits were with *Klosneuvirinae* subfamily, 11% with other members of the family *Mimiviridae*, 9% with Faustovirus, and 7% with Medusavirus, Orpheovirus and *Marseilleviridae* (**Figure 2B**). Regarding the best hits retrieved in Eukaryotes, the majority (30%) were with Metazoans, 22% with Fungi, and 12% with Viridiplantae. For the prokaryotes, most hits were with Bacteroidetes (≈23%), followed by Gammaproteobacteria (≈7%) and Firmicutes and Alphaproteobacteria (≈5% each).

Seven genes were found to carry introns. Five genes coding for DNA-directed RNA polymerase subunit 1 (RPB1), DNA-directed RNA polymerase subunit 2 (RPB2), ribonucleoside-diphosphate reductase large subunit, ankyrin repeat protein and DNA processing protein A (DprA) contained a single intron each. In contrast, the gene encoding a family B DNA polymerase elongation subunit contained two introns (**Supplementary Table 2**).

The ORF annotation indicated that 5.1% of the clandestinovirus genes correspond to the NCLDV core genes described by Koonin and Yutin (2010; **Table 2**), including four of the five inferred ancestral NCLDV proteins, namely, major capsid protein, the family B DNA polymerase, the A32-like DNA-packaging ATPase and the viral late transcription factor

TABLE 2 | Nucleo-cytoplasmic large DNA viruses (NCLDV) core genes in the clandestinovirus genome.

Clandestinovirus NCLDV core genes		
ORF n°	***NCLDV gene annotation***	***Associated functions***
203, 509	DNA or RNA helicase of superfamily II	DNA replication, recombination, and repair
315	D5-like helicase primase	DNA replication, recombination, and repair
211,212,213	DNA polymerase elongation subunit family B	DNA replication, recombination, and repair
299	ATP-dependent DNA ligase	DNA replication, recombination, and repair
582	FLAP-like endonuclease XPG	DNA replication, recombination, and repair
406	DNA topoisomerase II	DNA replication, recombination, and repair
49, 50	Ribonucleotide diphosphate reductase large subunit	Nucleotide metabolism
20	Ribonucleotide reductase small subunit	Nucleotide metabolism
524	Thymidylate synthase	Nucleotide metabolism
73, 389, 435	Dual specificity phosphatases; Ser/Thr, and Tyr proteins phosphatases	Other metabolic functions
269	Ubiquitin-conjugating enzyme 2	Other metabolic functions
445, 446	DNA-directed RNA polymerase subunit alpha	Transcription and RNA processing
25, 26	DNA-directed RNA polymerase subunit beta	Transcription and RNA processing
75	Divergent DNA-directed RNA polymerase subunit 5	Transcription and RNA processing
437	mRNA capping enzyme large subunit	Transcription and RNA processing
219, 573	Nudix hydrolase	Transcription and RNA processing
84	RNA ligase	Transcription and RNA processing
365	Transcription factor S-II (TFIIS)	Transcription and RNA processing
537	Poxvirus late transcription factor VLTF3 like	Transcription and RNA processing
364	Disulfide oxidoreductase; Erv1/Alr family	Virion structure and morphogenesis
154, 366	NCLDV major capsid protein	Virion structure and morphogenesis
13	A32-like packaging ATPase	Virion structure and morphogenesis

3 (VLTF3). In addition to the typical major capsid protein with the double jelly-roll fold, genes for the minor capsid proteins were found. They are homologous to the Paramecium bursaria chlorella virus 1 (PBCV-1) penton protein with the single jelly-roll fold found at the five-fold vertices of the icosahedral capsid as well as the PBCV-1 proteins P8 and P9, both involved in bridging and stabilizing the capsomers. This strongly suggests that the assembly principle of the clandestinovirus capsid is similar to that of phycodnaviruses and, likely, most other NCLDVs.

The clandestinovirus genome encodes several proteins involved in metabolic pathways. Nicotinamide/nicotinate mononucleotide adenylyl transferase (NMNAT) and NADAR (NAD^+ and ADP-ribose) participate in NAD^+ biosynthesis and NAD^+-utilization pathway, whereas cystathionine gamma-synthase and dihydrofolate reductase are involved in the synthesis of cystathionine and purines, respectively. Interestingly, we also identified a homolog of RNase T2, which has been recently described in Tupanvirus soda lake and discovered in Fadolivirus (Rolland et al., 2020; Schulz et al., 2020a).

We found that the virus encodes 11 serine/threonine kinases (**Supplementary Table 3**), including homologs of the cell cycle-related kinases, and three protein phosphatases, including mitogen-activated protein phosphatases. Remarkably, clandestinovirus encodes two homologs of Cyclin A2 (Cheng et al., 2006; Zhang et al., 2019), a factor that regulates cell cycle progression by interacting with two different cyclin-dependent kinases (CDK): CDK2 during S phase and CDK1 during the transition from G2 to M phase (Pagano et al., 1992). In addition, clandestinovirus encodes a homolog of Cdc123, an ATP-grasp fold protein involved in cell cycle control by promoting the transition from G1 to S phase (Panvert et al., 2015). Thus, clandestinovirus potentially orchestrate the cell cycle through controlling the phosphorylation state of key cellular factors as well as by direct interaction with the cellular or viral factors.

Furthermore, five LAGLIDADG endonuclease and two HNH family intron-encoded homing endonuclease were identified, which might have a potential impact on introns. Similar to phycodnaviruses, clandestinovirus encodes one DNA methyltransferase and eight restriction endonucleases of the HNH, PD-(D/E)XK and ββα-Me families which might be involved in the degradation of cellular DNA (Agarkova et al., 2006; Coy et al., 2020). In addition, the virus encodes a homolog of the mimivirus R354-like nuclease, a part of the MIMIVIRE system, with the closest homologs in phycodnaviruses (Levasseur et al., 2016). Moreover, several proteins with functions linked to the nucleus and DNA replication/transcription were identified. These include four BTB/POZ proteins (transcriptional regulators), several transcription factors as well as histone-like proteins.

Interestingly, the virus also encodes 10 proteins functioning in mitochondria (**Supplementary Table 4**), including three copies of the mitochondrial chaperone BCS1, which translocate substrates across the mitochondrial inner membrane without previous unfolding (Kater et al., 2020); mitochondrial deoxyguanosine kinase; Dynamin 1-like protein (DNM1-L),

FIGURE 3 | Phylogenetic tree of the DNA polymerase B sequences. The analysis was performed using the Maximum Likelihood method (ML) with a JTT substitution matrix in 1,000 replicates. Branch values lower than a bootstrap value of 0.5 were deleted. Colors were assigned for different group of viruses: blue for Mimiviruses and extended *Mimiviridae*; green for Pandoraviruses, Mollivirus sibericum, Emiliania huxleyi viruses (light green) and *Phycodnaviridae* (dark green); purple for groups of *Asfarviridae*, Faustoviruses, Pacmanvirus, and Kaumoabevirus; gray for Marseilleviridae; red for Orpheovirus, Solumvirus, Solivirus, Cedratviruses, and Pithovirus sibericum; and orange for *Iridoviridae*. Clandestinovirus is highlighted in bold and italic.

TABLE 3 | Description of proteins associated with histones in Clandestinovirus, Marseillevirus, Lausannevirus, and Acanthamoeba castellanii medusavirus.

Clandestinovirus		Marseillevirus marseillevirus		Lausannevirus		Acanthamoeba castellanii medusavirus	
Histones							
Protein ID	**Annotation**	**Protein ID**	**Annotation**	**Protein ID**	**Annotation**	**Protein ID**	**Annotation**
CV_ORF456	Histone H2B/H2A	YP_003407138.1	Histone H2B/H2A fusion protein	YP_004347348.1	Histone H2B/H2A fusion protein	BBI30201.1	Histone H2B
CV_ORF417	Histone H3	YP_003407137.1	Histone H3	YP_004347349.1	Histone H3-like protein	BBI30395.1	Histone H3
CV_ORF467	Histone H4					BBI30394.1	Histone H4
CV_ORF99	Linker histone H1/H5					BBI30246.1	Linker histone H1
		YP_003406909.1	Histone 2A-domain-containing protein	YP_004347019.1	Histone 2A-domain-containing protein	BBI30458.1	Histone H2A

a mechanochemical GTPase that induces membrane fission in mitochondria (Ford et al., 2011); mitochondrial sulfhydryl oxidase Erv1p; mitochondrial aspartate/glutamate carrier protein Aralar/Citrin; and two copies of the mitochondrial/bacterial CCA-adding enzyme.

Phylogenetic Relationship to Other Known Giant Viruses

A core and pan-genome analyses were carried out to determine the relationship of clandestinovirus to the previously characterized giant viruses (**Supplementary Table 5**). We considered a gene to belong to a cluster of orthologous genes (COG) if at least one reciprocal best hit was obtained between clandestinovirus and one virus included in the family tested. The largest number of conserved genes is shared with the klosneuviruses, with a total of 58 COGs, followed by the *Mimiviridae* (42 COG). Somewhat surprisingly, only 16 COGs are shared between clandestinovirus and medusavirus, despite the phylogenetic proximity of their core genes.

To determine the placement of clandestinovirus among other giant viruses, phylogenetic analysis was performed on nine conserved proteins. First, we chose three proteins considered to be ancestral to all NCLDVs and commonly used to determine the phylogenetic relationships among giant viruses (family B DNA polymerase, major capsid protein and VLTF-3). We added six more proteins commonly encoded in giant virus genomes (RPB1, RPB2, RPB5, and A32-like genome packaging ATPase as well as large and small subunits of the ribonucleoside-diphosphate reductase) (Koonin and Yutin, 2019). Phylogenetic analysis has shown that in most gene trees clandestinovirus forms a sister group to medusavirus (**Figure 3** and **Supplementary Figures 2–9**). Notably, however, the genome of clandestinovirus is 65% larger than that of medusavirus (581,987 versus 381,277 bp), indicating that the two viruses are only distantly related, consistent with a limited number of shared COGs (see above). Distant relationship is also supported by the observation that genes with introns are different between the two viruses, except for those encoding the family B DNA polymerase elongation subunit and the large subunit of the ribonucleoside-diphosphate reductase. Clandestinovirus was even more distantly related to Emiliania huxleyi viruses, Molliviruses, and Pandoraviruses, being placed at the root of branches including these viruses in five trees out of nine.

Deciphering Viral Histones and Associated Proteins in Clandestinovirus

Clandestinovirus carries homologs of the genes coding for four core histones (H3, H4, and H2B/H2A) and an additional gene for the linker histone H1/H5. The comparison with giant viruses encoding histones (Marseillevirus, Lausannevirus, and Medusavirus) showed that both Clandestinovirus and Medusavirus encode the four histones and a linker H1/H5, whereas Marseillevirus and Lausannevirus only encode three histones with a histone doublet H2B/H2A, a histone 2A-domain-containing protein and the histone H3 (**Table 3**).

To understand how clandestinovirus might benefit from encoding its own histones, we investigated the conservation of specific modification sites in N-terminal histone tails by aligning their sequences with sequences of histones from eukaryotes (**Figure 4** and **Supplementary Table 6**). The alignment of sequences for histone H2B and H4 showed two conserved sites of tail modification: the lysine 20 in position 3 of the H2B domain and lysine 5 in H4. For the histone domain H2A, lysines in positions 5 and 11 of the H2A domain from clandestinovirus coincide with lysines 9 and 15 of the consensus H2A sequence. Those two positions are the acetylation sites. The most conserved sites of modification are in the histone H3 with five sites of methylation on arginine 2 and lysines 4, 9, and 36. There are also five sites of acetylation (lysines 4, 9, 14, and 18) and one site of phosphorylation (threonine 11).

The phylogeny of histones indicates a branching of clandestinovirus between a clade including medusavirus and *Marseilleviridae* on one side and eukaryotic sequences on the other side (**Figure 5**). The phylogenetic distance between the histones of clandestinovirus and the host *V. vermiformis* suggested an acquisition from a source other than an amoebal host.

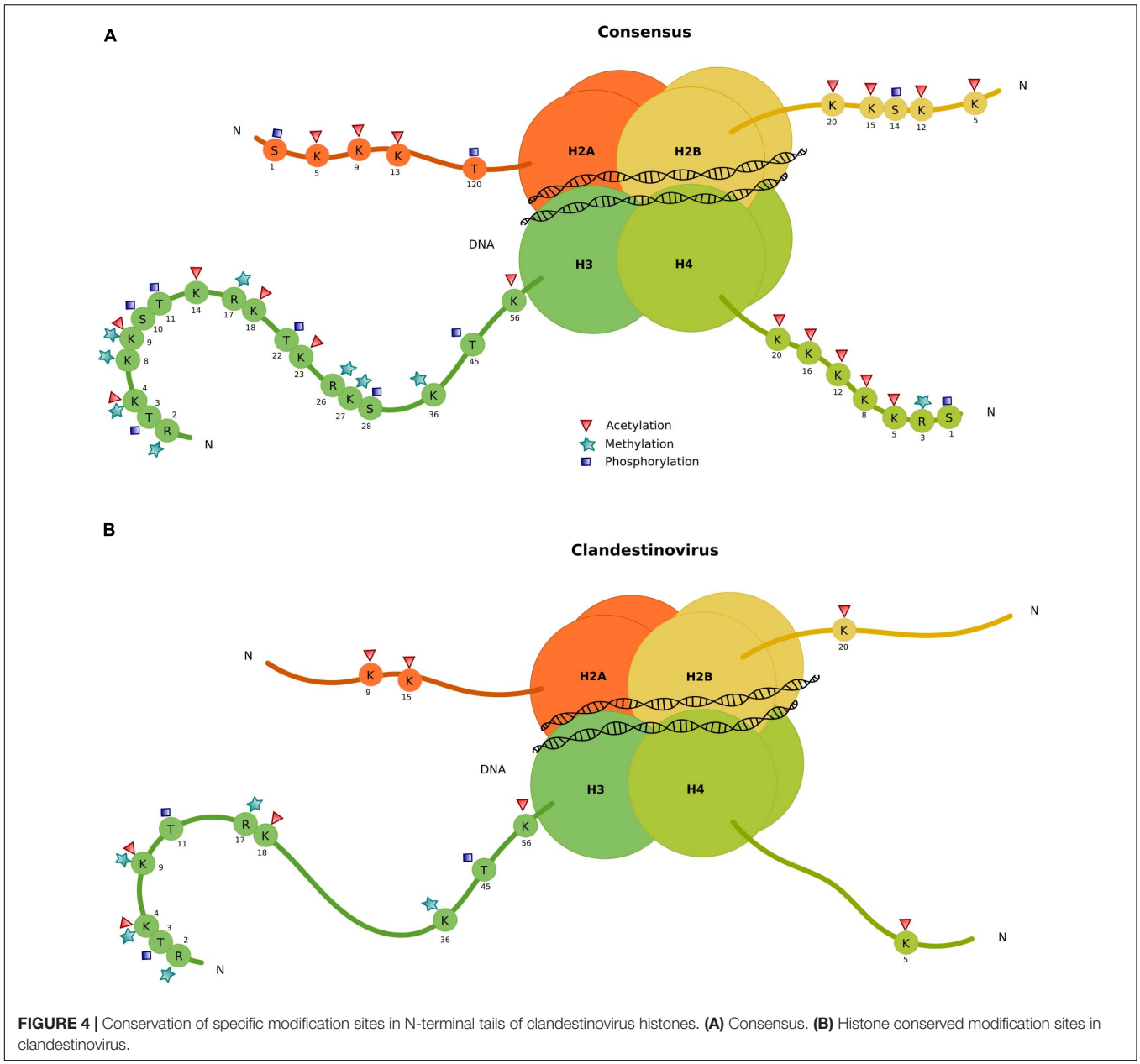


FIGURE 4 | Conservation of specific modification sites in N-terminal tails of clandestinovirus histones. **(A)** Consensus. **(B)** Histone conserved modification sites in clandestinovirus.

DISCUSSION

In this work, we described a novel giant virus isolated on *V. vermiformis* and named clandestinovirus ST1. It is a virus that was discovery in co-culture with another giant virus, faustovirus ST1 (Louazani et al., 2017). Surprisingly, clandestinovirus is not lytic and is released without causing a fast lysis of its host. It showed a viral cycle similar to that described for Acanthamoeba castellanii medusavirus, a giant virus recently discovered in Japan (Yoshikawa et al., 2019). Both viruses have an unusual replication cycle when compared to other giant viruses, with the entry into the nucleus of their amoebal hosts. However, unlike medusavirus, clandestinovirus replication, including virion assembly, takes place in the viral factories formed directly inside the nucleus.

Seven introns have been identified in the genome of clandestinovirus, as many as in medusavirus. In addition, seven homing endonucleases have been detected, five of the LAGLIDADG family and two of the HNH family. These enzymes bind and cleave DNA in specific patterns, providing the location for the insertion of intron-creating sequences if they are in coding regions (Stoddard, 2005), which might allow for the insertion of the introns.

Many viruses are known to manipulate the cell cycle to ensure the most favorable intracellular conditions for virus replication (Bagga and Bouchard, 2014; Fan et al., 2018; Liu J. et al., 2021). Interestingly, clandestinovirus encodes two cyclins with homologs in giant viruses and herpesviruses. Viral cyclins have been studied in the case of herpesviruses and were shown to

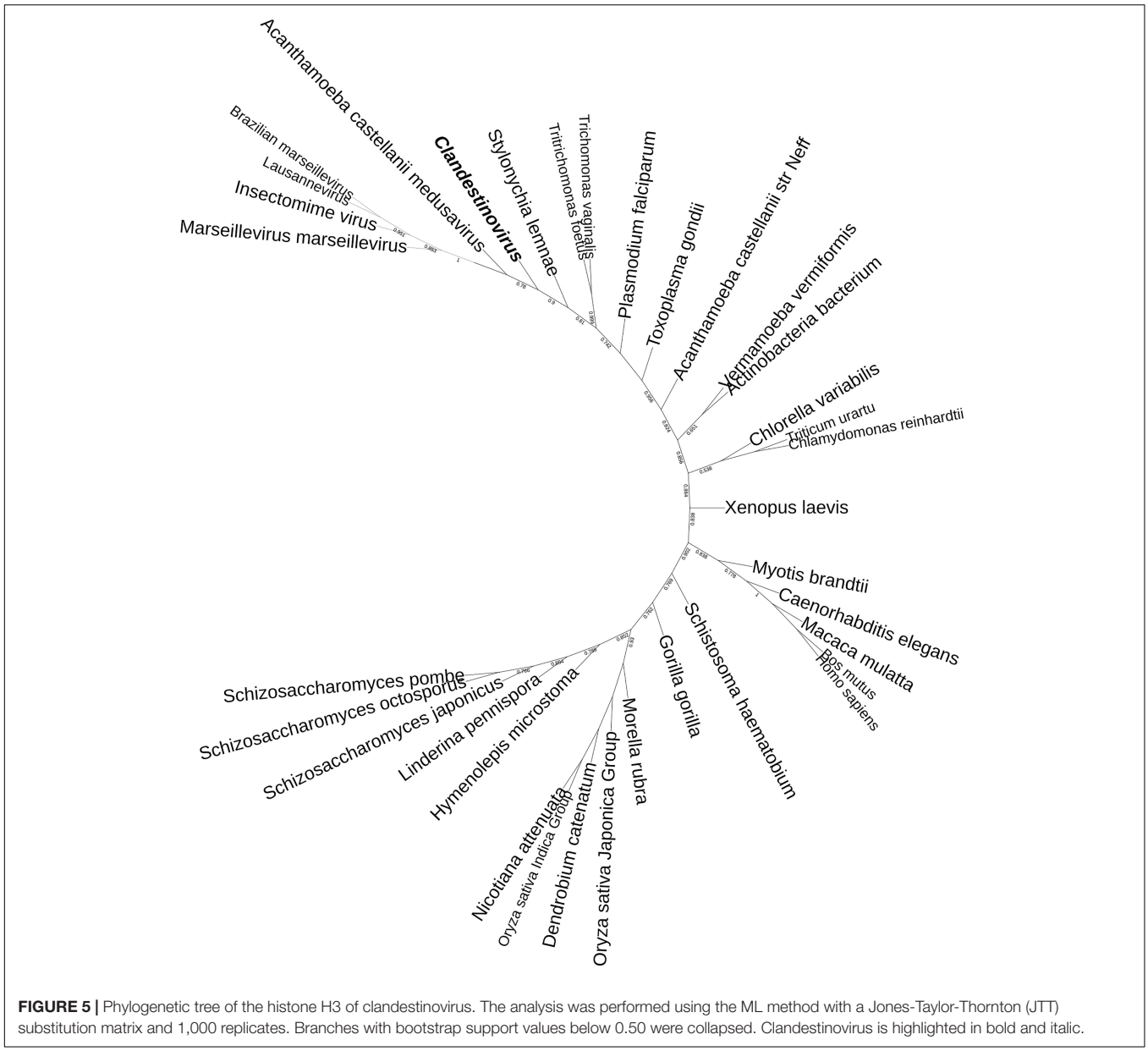


FIGURE 5 | Phylogenetic tree of the histone H3 of clandestinovirus. The analysis was performed using the ML method with a Jones-Taylor-Thornton (JTT) substitution matrix and 1,000 replicates. Branches with bootstrap support values below 0.50 were collapsed. Clandestinovirus is highlighted in bold and italic.

bind to one or more cellular cyclin-dependent kinases and phosphorylate an increased number of substrates compared with their cellular counterparts, thereby aiding virus replication (Hardwick, 2000; Ojala et al., 2000; Verschuren et al., 2004). Notably, clandestinovirus encodes multiple protein kinases and dephosphatases, which could further promote virus replication through orchestration of the cell cycle. Furthermore, the virus encodes several proteins functioning in mitochondria, among which mitochondrial chaperone BCS1 has been also detected in the medusaviruses but in single copy (Yoshikawa et al., 2019; Yoshida et al., 2021). The presence of mitochondria surrounding viral factories might indicate that these proteins help to stimulate some mitochondrial activities. Collectively, these observations suggest that clandestinovirus takes control of the cell cycle and optimizes the mitochondrial functions to boost its replication. Recent findings with Acanthamoeba castellanii medusavirus support this hypothesis (Zhang et al., 2021). Indeed, the authors have shown an alteration of the transcription of nuclear genes and a maintenance of the genes associated with mitochondria.

The number of COGs shared with the other giant virus families is inconsistent with the phylogeny of the core genes. Indeed, clandestinovirus shares the highest number of genes with klosneuviruses and tupanvirus, while in phylogenies, it forms a sister group to the medusavirus, despite disparity in genome sizes between the two viruses. One explanation is that the pan-genome of klosneusviruses is much larger than that of other virus groups and that overrepresentation of klosneusviruses in the database introduces a bias.

Furthermore, the annotation of clandestinovirus revealed the presence in its genome of numerous genes encoding proteins with functions associated or linked to the nucleus. Indeed, we found transcriptional factors, transcriptional regulators and proteins necessary for the viral DNA replication and transcription. Among them, we also retrieved homologs of the four histones, one of which is a histone doublet H2B/H2A, and the linker H1/H5. The presence of these proteins suggests that the virus might form nucleosomes, a complex associating DNA and histones. This complex is subject to post-translational modifications which either activate or repress the transcription of DNA necessary for viral replication (Kouzarides, 2007). In particular, these modifications include methylation and acetylation at particular residues located within histone tails and are well described in eukaryotic organisms. The analysis of these modification sites in clandestinovirus point to a conservation of these marks mostly on histone H3. They notably encompass acetylations, methylations or phosphorylations on well identified amino acids (Huang et al., 2014). In general, histone acetylation enables the activation of transcription, especially on lysine 56 of histone H3 (H3K56) (Tessarz and Kouzarides, 2014). A high level of acetylation combined with trimethylation of H3K4, H3K36, and H3K79 corresponds to actively transcribed euchromatin. In contrast, methylation sites on lysines H3K9, H3K27, and H4K20 are linked to the inhibition of transcription (Kouzarides, 2007). In clandestinovirus, only three activation sites (H3K4, H3K36, and H3K56) and one repressive site (H3K9) are conserved. Phylogenetic analysis of the viral histone places clandestinovirus at the base of the viral clade including medusavirus and marseilleviruses.

Moreover, histones are not the only elements associated with viral DNA replication found in clandestinovirus. Indeed, a potential restriction-modification system (R-M system) including the DNA *N*-6-adenine methyltransferase (Dam) and type-II restriction endonucleases were identified (eight HNH restriction endonucleases of different families). This system is linked to DNA methylation and is widely found in giant viruses (Jeudy et al., 2020). Some authors suggest that it could be a protection system against the host or other microorganisms (viruses or bacteria). In chloroviruses, the R-M system is involved in the degradation of cellular DNA to allow viruses to use the nucleotide pools created (Agarkova et al., 2006). In the case of clandestinovirus, the R-M system might be either a protection system that make it possible to coexist with faustovirus ST1 and a weapon to allow it to have access to more nucleotides for its replication.

Clandestinovirus is a new giant virus discovered in *V. vermiformis*. Its replicative cycle takes place in the host nucleus, and its genomic characteristics and phylogeny suggest that it is a distant relative of Acanthamoeba castellanii medusavirus. We suggest that clandestinovirus represents a separate genus within the family *Medusaviridae* proposed by Yoshikawa et al. (2019). Clandestinovirus opens new perspectives for this family, broadening the host range to a new amoeba species. Further investigations on the diversity of the *Medusaviridae* family should expand our knowledge on the co-evolution between large DNA viruses and eukaryotes.

DATA AVAILABILITY STATEMENT

The data presented in the study are deposited in the NCBI database repository under accession numbers MZ420154 and MZ420155.

AUTHOR CONTRIBUTIONS

AL and BL designed and supervised the study. CR, JA, DS-B, MK, and AL performed sample collection, virus isolation, experiments, and analyses. CR, JA, MK, and AL wrote the manuscript. All authors read and approved the final version of the manuscript.

FUNDING

This research was funded by the Mediterranee-Infection Foundation and by the French National Research Agency under the program "Investissements d'Avenir," reference ANR-10-IAHU-03. This research was supported by a grant from the Institut Universitaire de France (IUF, Paris, France) awarded to AL.

ACKNOWLEDGMENTS

The authors acknowledge Fabrizio Di Pinto for the technical support for the electron microscopy.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2021.715608/full#supplementary-material

REFERENCES

Abrahão, J., Silva, L., Silva, L. S., Khalil, J. Y. B., Rodrigues, R., Arantes, T., et al. (2018). Tailed giant Tupanvirus possesses the most complete translational apparatus of the known virosphere. *Nat. Commun.* 9:749. doi: 10.1038/s41467-018-03168-1

Agarkova, I. V., Dunigan, D. D., and Etten, J. L. V. (2006). Virion-Associated Restriction Endonucleases of Chloroviruses. *J. Virol.* 80, 8114–8123. doi: 10.1128/JVI.00486-06

Andreani, J., Aherfi, S., Bou Khalil, J. Y., Di Pinto, F., Bitam, I., Raoult, D., et al. (2016). Cedratvirus, a Double-Cork Structured Giant Virus, is a Distant Relative of Pithoviruses. *Viruses* 8:300. doi: 10.3390/v8110300

Antipov, D., Korobeynikov, A., McLean, J. S., and Pevzner, P. A. (2016). hybridSPAdes: an algorithm for hybrid assembly of short and long reads. *Bioinformatics* 32, 1009–1015. doi: 10.1093/bioinformatics/btv688

Bagga, S., and Bouchard, M. J. (2014). "Cell Cycle Regulation During Viral Infection," in *Cell Cycle Control: Mechanisms and Protocols* Methods in

Molecular Biology, eds E. Noguchi and M. C. Gadaleta (New York, NY: Springer), 165–227. doi: 10.1007/978-1-4939-0888-2_10

Besemer, J., Lomsadze, A., and Borodovsky, M. (2001). GeneMarkS: a self-training method for prediction of gene starts in microbial genomes. Implications for finding sequence motifs in regulatory regions. *Nucleic Acids Res.* 29, 2607–2618.

Bou Khalil, J. Y., Andreani, J., Raoult, D., and La Scola, B. (2016). A Rapid Strategy for the Isolation of New Faustoviruses from Environmental Samples Using Vermamoeba vermiformis. *J. Vis. Exp. JoVE* 2016:54104. doi: 10.3791/54104

Boyer, M., Yutin, N., Pagnier, I., Barrassi, L., Fournous, G., Espinosa, L., et al. (2009). Giant Marseillevirus highlights the role of amoebae as a melting pot in emergence of chimeric microorganisms. *Proc. Natl. Acad. Sci.* 106, 21848–21853. doi: 10.1073/pnas.0911354106

Cheng, K.-Y., Noble, M. E. M., Skamnaki, V., Brown, N. R., Lowe, E. D., Kontogiannis, L., et al. (2006). The Role of the Phospho-CDK2/Cyclin A Recruitment Site in Substrate Recognition. *J. Biol. Chem.* 281, 23167–23179. doi: 10.1074/jbc.M600480200

Colson, P., Yutin, N., Shabalina, S. A., Robert, C., Fournous, G., La Scola, B., et al. (2011). Viruses with More Than 1,000 Genes: Mamavirus, a New Acanthamoeba polyphagamimivirus Strain, and Reannotation of Mimivirus Genes. *Genome Biol. Evol.* 3, 737–742. doi: 10.1093/gbe/evr048

Coy, S. R., Gann, E. R., Papoulis, S. E., Holder, M. E., Ajami, N. J., Petrosino, J. F., et al. (2020). SMRT Sequencing of Paramecium Bursaria Chlorella Virus-1 Reveals Diverse Methylation Stability in Adenines Targeted by Restriction Modification Systems. *Front. Microbiol.* 11:887. doi: 10.3389/fmicb.2020.00887

Edgar, R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 32, 1792–1797. doi: 10.1093/nar/gkh340

Fan, Y., Sanyal, S., and Bruzzone, R. (2018). Breaking Bad: How Viruses Subvert the Cell Cycle. *Front. Cell. Infect. Microbiol.* 8:396. doi: 10.3389/fcimb.2018.00396

Ford, M. G. J., Jenni, S., and Nunnari, J. (2011). The crystal structure of dynamin. *Nature* 477, 561–566. doi: 10.1038/nature10441

Gong, Z., Zhang, Y., and Han, G.-Z. (2020). Molecular fossils reveal ancient associations of dsDNA viruses with several phyla of fungi. *Virus Evol.* 6:8. doi: 10.1093/ve/veaa008

Hardwick, J. M. (2000). Cyclin' on the viral path to destruction. *Nat. Cell Biol.* 2, E203–E204. doi: 10.1038/35041126

Huang, H., Sabari, B. R., Garcia, B. A., Allis, C. D., and Zhao, Y. (2014). *Cell* 159, 458.e–458.e. doi: 10.1016/j.cell.2014.09.037

Hyatt, D., Chen, G.-L., LoCascio, P. F., Land, M. L., Larimer, F. W., and Hauser, L. J. (2010). Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* 11:119. doi: 10.1186/1471-2105-11-119

Iyer, L. M., Aravind, L., and Koonin, E. V. (2001). Common Origin of Four Diverse Families of Large Eukaryotic DNA Viruses. *J. Virol.* 75, 11720–11734. doi: 10.1128/JVI.75.23.11720-11734.2001

Jeudy, S., Rigou, S., Alempic, J.-M., Claverie, J.-M., Abergel, C., and Legendre, M. (2020). The DNA methylation landscape of giant viruses. *Nat. Commun.* 11:2657. doi: 10.1038/s41467-020-16414-2

Kater, L., Wagener, N., Berninghausen, O., Becker, T., Neupert, W., and Beckmann, R. (2020). Structure of the Bcs1 AAA-ATPase suggests an airlock-like translocation mechanism for folded proteins. *Nat. Struct. Mol. Biol.* 27, 142–149. doi: 10.1038/s41594-019-0364-1

Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N., and Sternberg, M. J. E. (2015). The Phyre2 web portal for protein modeling, prediction and analysis. *Nat. Protoc.* 10, 845–858. doi: 10.1038/nprot.2015.053

Koonin, E. V., Dolja, V. V., Krupovic, M., Varsani, A., Wolf, Y. I., Yutin, N., et al. (2020). Global Organization and Proposed Megataxonomy of the Virus World. *Microbiol. Mol. Biol. Rev.* 84:19. doi: 10.1128/MMBR.00061-19

Koonin, E. V., and Yutin, N. (2010). Origin and Evolution of Eukaryotic Large Nucleo-Cytoplasmic DNA Viruses. *Intervirology* 53, 284–292. doi: 10.1159/000312913

Koonin, E. V., and Yutin, N. (2019). "Chapter Five - Evolution of the Large Nucleocytoplasmic DNA Viruses of Eukaryotes and Convergent Origins of Viral Gigantism," in *Advances in Virus Research*, eds M. Kielian, T. C. Mettenleiter, and M. J. Roossinck (Cambridge, MA: Academic Press), 167–202. doi: 10.1016/bs.aivir.2018.09.002

Kouzarides, T. (2007). Chromatin Modifications and Their Function. *Cell* 128, 693–705. doi: 10.1016/j.cell.2007.02.005

La Scola, B., Audic, S., Robert, C., Jungang, L., De Lamballerie, X., Drancourt, M., et al. (2003). A giant virus in amoebae. *Science* 299:2033. doi: 10.1126/science.1081867

Lechner, M., Findeiß, S., Steiner, L., Marz, M., Stadler, P. F., and Prohaska, S. J. (2011). Proteinortho: Detection of (Co-)orthologs in large-scale analysis. *BMC Bioinformatics* 12:124. doi: 10.1186/1471-2105-12-124

Letunic, I., and Bork, P. (2021). Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. *Nucleic Acids Res.* 2021:301. doi: 10.1093/nar/gkab301

Levasseur, A., Bekliz, M., Chabrière, E., Pontarotti, P., La Scola, B., and Raoult, D. (2016). MIMIVIRE is a defence system in mimivirus that confers resistance to virophage. *Nature* 531, 249–252. doi: 10.1038/nature17146

Liu, J., Cvirkaite-Krupovic, V., Baquero, D. P., Yang, Y., Zhang, Q., Shen, Y., et al. (2021). Virus-induced cell gigantism and asymmetric cell division in archaea. *Proc. Natl. Acad. Sci.* 2021:118. doi: 10.1073/pnas.2022578118

Liu, Y., Toner, C. M., Philippe, N., Jeudy, S., Zhou, K., Bowerman, S., et al. (2021). Melbournevirus-encoded histone doublets are recruited to virus particles and form destabilized nucleosome-like structures. *bioRxiv* 2021:441998. doi: 10.1101/2021.04.29.441998

Louazani, A. C., Andreani, J., Ouarhache, M., Aherfi, S., Baptiste, E., Levasseur, A., et al. (2017). Genome Sequences of New Faustovirus Strains ST1 and LC9, Isolated from the South of France. *Genome Announc* 5, e613–e617. doi: 10.1128/genomeA.00613-17

Marchler-Bauer, A., Bo, Y., Han, L., He, J., Lanczycki, C. J., Lu, S., et al. (2017). CDD/SPARCLE: functional classification of proteins via subfamily domain architectures. *Nucleic Acids Res.* 45, D200–D203. doi: 10.1093/nar/gkw1129

Mitchell, A. L., Attwood, T. K., Babbitt, P. C., Blum, M., Bork, P., Bridge, A., et al. (2019). InterPro in 2019: improving coverage, classification and access to protein sequence annotations. *Nucleic Acids Res.* 47, D351–D360. doi: 10.1093/nar/gky1100

Moniruzzaman, M., Martinez-Gutierrez, C. A., Weinheimer, A. R., and Aylward, F. O. (2020a). Dynamic genome evolution and complex virocell metabolism of globally-distributed giant viruses. *Nat. Commun.* 11, 1–11. doi: 10.1038/s41467-020-15507-2

Moniruzzaman, M., Weinheimer, A. R., Martinez-Gutierrez, C. A., and Aylward, F. O. (2020b). Widespread endogenization of giant viruses shapes genomes of green algae. *Nature* 2020, 1–5. doi: 10.1038/s41586-020-2924-2

Needham, D. M., Poirier, C., Hehenberger, E., Jiménez, V., Swalwell, J. E., Santoro, A. E., et al. (2019a). Targeted metagenomic recovery of four divergent viruses reveals shared and distinctive characteristics of giant viruses of marine eukaryotes. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* 374:20190086. doi: 10.1098/rstb.2019.0086

Needham, D. M., Yoshizawa, S., Hosaka, T., Poirier, C., Choi, C. J., Hehenberger, E., et al. (2019b). A distinct lineage of giant viruses brings a rhodopsin photosystem to unicellular marine predators. *Proc. Natl. Acad. Sci.* 116, 20574–20583. doi: 10.1073/pnas.1907517116

Ojala, P. M., Yamamoto, K., Castaños-Vélez, E., Biberfeld, P., Korsmeyer, S. J., and Mäkelä, T. P. (2000). The apoptotic v-cyclin-CDK6 complex phosphorylates and inactivates Bcl-2. *Nat. Cell Biol.* 2, 819–825. doi: 10.1038/35041064

Pagano, M., Pepperkok, R., Verde, F., Ansorge, W., and Draetta, G. (1992). Cyclin A is required at two points in the human cell cycle. *EMBO J.* 11, 961–971. doi: 10.1002/j.1460-2075.1992.tb05135.x

Panvert, M., Dubiez, E., Arnold, L., Perez, J., Mechulam, Y., Seufert, W., et al. (2015). Cdc123, a Cell Cycle Regulator Needed for eIF2 Assembly, Is an ATP-Grasp Protein with Unique Features. *Struct. Lond. Engl.* 1993, 1596–1608. doi: 10.1016/j.str.2015.06.014

Philippe, N., Legendre, M., Doutre, G., Couté, Y., Poirot, O., Lescot, M., et al. (2013). Pandoraviruses: Amoeba Viruses with Genomes Up to 2.5 Mb Reaching That of Parasitic Eukaryotes. *Science* 341:281. doi: 10.1126/science.1239181

Price, M. N., Dehal, P. S., and Arkin, A. P. (2010). FastTree 2 – Approximately Maximum-Likelihood Trees for Large Alignments. *PLoS One* 5:e9490. doi: 10.1371/journal.pone.0009490

Raoult, D., Audic, S., Robert, C., Abergel, C., Renesto, P., Ogata, H., et al. (2004). The 1.2-Megabase Genome Sequence of Mimivirus. *Science* 306:1344. doi: 10.1126/science.1101485

Rolland, C., Andreani, J., Cherif Louazani, A., Aherfi, S., Francis, R., Rodrigues, R., et al. (2019). Discovery and Further Studies on Giant Viruses at the IHU

Mediterranee Infection That Modified the Perception of the Virosphere. *Viruses* 11:312. doi: 10.3390/v11040312

Rolland, C., La Scola, B., and Levasseur, A. (2020). How Tupanvirus Degrades the Ribosomal RNA of Its Amoebal Host? The Ribonuclease T2 Track. *Front. Microbiol.* 11:1691. doi: 10.3389/fmicb.2020.01691

Sahmi-Bounsiar, D., Boratto, P. V., de, M., Oliveira, G. P., Khalil, J. Y. B., Scola, B. L., et al. (2019). Single Cell Micro-aspiration as an Alternative Strategy to Fluorescence-activated Cell Sorting for Giant Virus Mixture Separation. *JoVE J. Vis. Exp.* 2019:e60148. doi: 10.3791/60148

Schulz, F., Andreani, J., Francis, R., Boudjemaa, H., Khalil, J. Y. B., Lee, J., et al. (2020a). Advantages and Limits of Metagenomic Assembly and Binning of a Giant Virus. *mSystems* 5:20. doi: 10.1128/mSystems.00048-20

Schulz, F., Roux, S., Paez-Espino, D., Jungbluth, S., Walsh, D. A., Denef, V. J., et al. (2020b). Giant virus diversity and host interactions through global metagenomics. *Nature* 578, 432–436. doi: 10.1038/s41586-020-1957-x

Schulz, F., Yutin, N., Ivanova, N. N., Ortega, D. R., Lee, T. K., Vierheilig, J., et al. (2017). Giant viruses with an expanded complement of translation system components. *Science* 356:82. doi: 10.1126/science.aal4657

Steinegger, M., Meier, M., Mirdita, M., Vöhringer, H., Haunsberger, S. J., and Söding, J. (2019). HH-suite3 for fast remote homology detection and deep protein annotation. *BMC Bioinformatics* 20:473. doi: 10.1186/s12859-019-3019-7

Stoddard, B. L. (2005). Homing endonuclease structure and function. *Q. Rev. Biophys.* 38, 49–95. doi: 10.1017/S0033583505004063

Tessarz, P., and Kouzarides, T. (2014). Histone core modifications regulating nucleosome structure and dynamics. *Nat. Rev. Mol. Cell Biol.* 15, 703–708. doi: 10.1038/nrm3890

Thomas, V., Bertelli, C., Collyn, F., Casson, N., Telenti, A., Goesmann, A., et al. (2011). Lausannevirus, a giant amoebal virus encoding histone doublets. *Environ. Microbiol.* 13, 1454–1466. doi: 10.1111/j.1462-2920.2011.02446.x

Valencia-Sánchez, M. I., Abini-Agbomson, S., Wang, M., Lee, R., Vasilyev, N., Zhang, J., et al. (2021). The structure of a virus-encoded nucleosome. *Nat. Struct. Mol. Biol.* 28, 413–417. doi: 10.1038/s41594-021-00585-7

Verschuren, E. W., Jones, N., and Evan, G. I. Y. (2004). The cell cycle and how it is steered by Kaposi's sarcoma-associated herpesvirus cyclin. *J. Gen. Virol.* 85, 1347–1361. doi: 10.1099/vir.0.79812-0

Walker, P. J., Siddell, S. G., Lefkowitz, E. J., Mushegian, A. R., Dempsey, D. M., Dutilh, B. E., et al. (2019). Changes to virus taxonomy and the International Code of Virus Classification and Nomenclature ratified by the International Committee on Taxonomy of Viruses (2019). *Arch. Virol.* 164, 2417–2429. doi: 10.1007/s00705-019-04306-w

Yoshida, K., Zhang, R., Garcia, K. G., Endo, H., Gotoh, Y., Hayashi, T., et al. (2021). Draft Genome Sequence of Medusavirus Stheno, Isolated from the Tatakai River of Uji, Japan. *Microbiol. Resour. Announc.* 10:20. doi: 10.1128/MRA.01323-20

Yoshikawa, G., Blanc-Mathieu, R., Song, C., Kayama, Y., Mochizuki, T., Murata, K., et al. (2019). Medusavirus, a novel large DNA virus discovered from hot spring water. *J. Virol.* 2019:18. doi: 10.1128/JVI.02130-18

Yutin, N., and Koonin, E. V. (2012). Proteorhodopsin genes in giant viruses. *Biol. Direct* 7:34. doi: 10.1186/1745-6150-7-34

Zhang, R., Endo, H., Takemura, M., and Ogata, H. (2021). RNA-seq of the medusavirus suggests remodeling of the host nuclear environment at an early infection stage. *bioRxiv* 2021:439121. doi: 10.1101/2021.04.10.439121

Zhang, S., Tischer, T., and Barford, D. (2019). Cyclin A2 degradation during the spindle assembly checkpoint requires multiple binding modes to the APC/C. *Nat. Commun.* 10:3863. doi: 10.1038/s41467-019-11833-2

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Microbiology

ORIGINAL RESEARCH
published: 21 September 2021
doi: 10.3389/fmicb.2021.719703

Morphological and Genomic Features of the New Klosneuvirinae Isolate Fadolivirus IHUMI-VV54

Julien Andreani[1]†, Frederik Schulz[2†], Fabrizio Di Pinto[1], Anthony Levasseur[1], Tanja Woyke[2] and Bernard La Scola[1]**

[1]Aix-Marseille Université, IRD, APHM, MEPHI, IHU Méditerranée Infection, Marseille, France, [2]DOE Joint Genome Institute, Lawrence Berkeley National Laboratory, Berkeley, CA, United States

Since the discovery of *Mimivirus*, viruses with large genomes encoding components of the translation machinery and other cellular processes have been described as belonging to the nucleocytoplasmic large DNA viruses. Recently, genome-resolved metagenomics led to the discovery of more than 40 viruses that have been grouped together in a proposed viral subfamily named *Klosneuvirinae*. Members of this group had genomes of up to 2.4 Mb in size and featured an expanded array of translation system genes. Yet, despite the large diversity of the Klosneuvirinae in metagenomic data, there are currently only two isolates available. Here, we report the isolation of a novel giant virus known as Fadolivirus from an Algerian sewage site and provide morphological data throughout its replication cycle in amoeba and a detailed genomic characterization. The Fadolivirus genome, which is more than 1.5 Mb in size, encodes 1,452 predicted proteins and phylogenetic analyses place this viral isolate as a near relative of the metagenome assembled Klosneuvirus and Indivirus. The genome encodes for 66 tRNAs, 23 aminoacyl-tRNA synthetases and a wide range of transcription factors, surpassing Klosneuvirus and other giant viruses. The Fadolivirus genome also encodes putative vacuolar-type proton pumps with the domains D and A, potentially constituting a virus-derived system for energy generation. The successful isolation of Fadolivirus will enable future hypothesis-driven experimental studies providing deeper insights into the biology of the *Klosneuvirinae*.

Keywords: *Vermamoeba vermiformis*, giant virus, translation components, Mimiviridae, ATPase subunits, Klosneuvirinae, Fadolivirus

OPEN ACCESS

Edited by:
Frank O'Neill Aylward, Virginia Tech, United States

Reviewed by:
Masaharu Takemura, Tokyo University of Science, Japan
Felipe Hernandes Coutinho, Consejo Superior de Investigaciones Científicas (CSIC), Spain

****Correspondence:***
Bernard La Scola
bernard.la-scola@univ-amu.fr
Julien Andreani
miaguiabidou@gmail.com

†These authors have contributed equally to this work and share first authorship

Specialty section:
This article was submitted to Virology, a section of the journal Frontiers in Microbiology

Received: *02 June 2021*
Accepted: *18 August 2021*
Published: *21 September 2021*

Citation:
Andreani J, Schulz F, Di Pinto F, Levasseur A, Woyke T and La Scola B (2021) Morphological and Genomic Features of the New Klosneuvirinae Isolate Fadolivirus IHUMI-VV54. Front. Microbiol. 12:719703. doi: 10.3389/fmicb.2021.719703

INTRODUCTION

In 2003, with the first description of Mimivirus (La Scola, 2003), viruses were revealed to be more complex entities than previously thought and with extraordinary properties, e.g., numerous tRNAs, aminoacyl-tRNA synthetases (aaRS) and themselves being infected by virophages (La Scola et al., 2008; Gaia et al., 2014). In terms of capsid size, as well as gene content, this first discovery created a paradigm shift and shattered extant viral definition (Forterre, 2016; Colson et al., 2017). 18 years later, giant viruses with ovoid forms, even larger genome sizes, and viruses with no capsid have been discovered with the successive

isolations of Pandoraviruses (Philippe et al., 2013; Antwerpen et al., 2015; Aherfi et al., 2018; Legendre et al., 2018), Pithoviruses (Legendre et al., 2014; Levasseur et al., 2016), Cedratviruses (Andreani et al., 2016; Bertelli et al., 2017; Silva et al., 2018; Rodrigues et al., 2018a; Boudjemaa et al., 2020), and Orpheovirus (Andreani et al., 2018). In addition, the description of Meelsvirus revealed that these potential viruses can take extremely varied forms and are associated with a complex formation of their capsids (Shinn and Bullard, 2018). These large viruses share a common origin and a novel name has been proposed: Nucleocytoviricota (replacing the term nucleocytoplasmic large DNA viruses (NCLDV; Koonin et al., 2020).

In parallel, advances in (meta) genomic sequencing of viruses expanded rapidly (Verneau et al., 2016; Schulz et al., 2017, 2018, 2020; Wilson et al., 2017; Moniruzzaman et al., 2020a,b). In 2017, a metagenomic study reported the detection and genome reconstruction of four new viruses affiliated with the Mimiviridae which encoded expanded protein biosynthesis components, Klosneuvirus, Indivirus, Catovirus, and Hokovirus, proposed as a sub-family, *Klosneuvirinae* (Schulz et al., 2017). Just 1 year later, the diversity of this unexplored viral group was further expanded with the discovery of seven additional metagenome assembled viruses from soil (Schulz et al., 2018) and the first isolate affiliated with the Klosneuvirinae, Bodo saltans virus (Deeg et al., 2018). In contrast to many other members of the Klosneuvirinae, Bodo saltans virus encodes for only two functional aminoacyl-tRNA synthetases (aaRS) and no tRNA. In the meantime, two novel members of the Mimiviridae were isolated, Tupanvirus soda lake and Tupanvirus deep ocean, possessing 70 tRNA and 20 aaRS, and similar and slightly more complete translational components than the metagenomic Klosneuvirus (Abrahão et al., 2018). Tupanviruses do not group with the Klosneuvirinae but with the Megamimivirinae, sharing a hypothetical ancestor with the three lineages of Mimiviruses A, B, and C (Rodrigues et al., 2018b). The first isolate of the Klosneuvirinae that had an expanded set of translation system genes was the icosahedral virus in *Vermamoeba vermiformis,* named Yasminevirus (Bajrai et al., 2019). Its genome is one of the largest in the Mimiviridae with 2.1 Mb and possesses various components of a complete translation system. The discovery of these viruses has opened up new perspectives regarding their evolution in relation to eukaryotes. Recently, metagenomic has revealed the widespread presence of energy metabolism encoded in the Nucleocytoviricota, such as photosynthesis genes and enzymes playing roles in glycolysis among others (Schulz et al., 2020; Moniruzzaman et al., 2020a).

In this study, we report the isolation and detailed description of the closest relative of the metagenomic Klosneuvirus from an Algerian sewage sample, representing the third viral isolate of the *Klosneuvirinae*. Our study provides information on the viral replicative cycle and host range, as well as its phylogeny and genomic features, revealing a more complete picture of the *Klosneuvirinae.* Our isolate contributes toward establishing an experimental framework of viral strains that may be used to study the biology of this diverse subfamily of the Mimiviridae.

MATERIALS AND METHODS

Sample Collection and Virus Isolation

Samples from which the virus was isolated were part of a study aiming to evaluate giant virus diversity in north-west Algeria (Boudjemaa et al., 2020). Briefly, a sample was collected from sewage in Sidi Bel Abbès (Algeria) on the September 1, 2017. As previously reported (Geballa-Koukoulas et al., 2020, 2021), we used *Vermamoeba vermiformis* stain CDC19 as cell support. The amoebae were harvested after 48 h of culture in peptone yeast extract glucose medium (Eurobio, France) when a concentration of 1.10^6 amoebae/mL was reached. Cells were then rinsed twice in homemade Page's Amoeba Saline and pelleted at 700 × g for 10 min. The amoebae were then resuspended in the starvation medium (Reteno et al., 2015). A cell suspension of 500 µl per well was distributed onto a 24-well plate. The samples were vortexed, and 50 µl was added to each well. Remaining wells served as negative controls. The plate was incubated at 30°C for 4 days in order to monitor any potential cytopathic effect. This co-culture was repeated twice in the same order.

Host Range Survey

Multiple hosts were tested by inoculation of Fadolivirus on *Vermamoeba vermiformis* (ATCC® 50237, initial support of isolation), *Vermamoeba vermiformis* strain chuv172, *Acanthamoeba castellanii* strain Neff (ATCC® 30010), *Acanthamoeba polyphaga* strain Linc-AP1, *Acanthamoeba royreba* strain Oak ridge (ATCC® 30884), *Disctyostelium discoideum* (ATCC® 44841), the flagellated alga *Tetraselmis* sp. (IHU isolate), and the protist *Cafeteria roenbergensis* (IHU isolate). The lytic events were monitored by observations under an inverted microscope, as described for Tupanvirus (Abrahão et al., 2018), for 5 days.

Flow Cytometry Analyses

According to previously described protocols (Khalil et al., 2016; Andreani et al., 2017), we centrifuged the viral supernatant at 700 × *g* for 10 min, added a 1:100 SYBR green dye dilution (SYBR green I nucleic acid gel stain; Molecular Probes, Life Technologies), and then heated the mixture at 80°C for 3 min. Stained viral populations were acquired on a BD LSR Fortessa (BD Biosciences) cytometer and compared with previously recorded gates for known viral isolates on FlowJo software.

Electron Microscopy Observations

We used protocols similar to previous studies (Andreani et al., 2016, 2017) for embedding and transmission electron microscopy. The only difference was the infection time points at which the samples were collected. The time points were 0, 4, 8, 12, 16, 20, 24, and 36 h post-infection. Microscopy was performed using a Tecnai G20 electron microscope (FEI, Germany) operating at 200 keV, and the size of the particles was measured using ImageJ.[1]

[1]https://imagej.nih.gov/ij/

Viral Production and Purity Control

We successively inoculated diluted viral supernatant on *V. vermiformis* at a dilution factor of 10 in order to clone the virus before its production. End point dilution was assessed for 5 days and the lysis was controlled by inverted microscopy. For the production and purification processes, 15 flasks of 150 cm^2 (Corning®, Corning, NY, United States) were infected. After 4 days, we filtrated all the supernatants from flasks using 0.8 µm filters. The filtrate supernatant was then pelleted using the Beckman coulter® Optima™ XPN-80 - IVD (Beckman, France) at 50,000 × g for 30 min. A 25% sucrose gradient was used for the final purification step. After finalizing production, we extracted DNA by using the EZ1 Advanced XL device (Qiagen, Courtaboeuf, France).

Genome Sequencing and Genome Assembly

Genomic DNA of Fadolivirus was sequenced on the MiSeq platform (Illumina Inc., San Diego, CA, United States) using a paired end strategy. In order to pool the sample with 18 other genomic projects for sequencing, a barcoded library was created using the Nextera XT DNA sample prep kit (Illumina). In detail, Fadolivirus gDNA was quantified using a Qubit assay using the high sensitivity kit (Life technologies, Carlsbad, CA, United States). DNA concentration was 5.7 ng/µl. To prepare the paired end library, the gDNA was diluted to provide 1 ng of input. The "tagmentation" step fragmented and tagged the DNA. A limited cycle PCR amplification (12 cycles) completed the tag adapters and introduced dual-index barcodes. After purification on AMPure XP beads (Beckman Coulter Inc., Fullerton, CA, United States, the libraries were normalized on specific beads according to the Nextera XT protocol (Illumina). Normalized barcoded libraries were then pooled into a single library for sequencing on the MiSeq. The pooled single strand library was loaded onto the reagent cartridge and the instrument along with the flow cell. Automated cluster generation and paired end sequencing with dual index reads were performed in a single 39-h run in 2 × 250-bp. A total of 8.5Gb of Fadolivirus genomic sequence was obtained from a 927,000 cluster density per mm^2 with a cluster passing quality control filters of 94.1% (16,738,000 clusters). Within this run, the index representation for this virus was determined to represent 6.8%. To improve the Fadolivirus assembly, a MinIon Oxford Nanopore run was performed.

The Oxford Nanopore approach was performed on 1D genomic DNA sequencing using the MinIon device and SQK-LSK108 kit. The sequencing library was constructed using 1.5 µg genomic DNA previously extracted without fragmentation and end repair. Adapters were ligated to both ends of genomic DNA. After purification on AMPure XP beads (Beckman Coulter Inc., Fullerton, CA, United States), the library was quantified by a Qubit assay with the high sensitivity kit (Life technologies, Carlsbad, CA, United States). Approximately 20 ng of the tethered library was loaded on the flow cell *via* the SpotON port. 317 active pores were detected for sequencing and the WIMP workflow was chosen for live bioinformatics analysis. After 36 h of run time and end life of the flow cell, 13,136 raw reads were generated. The software EPI2ME lead to 13,122 analyzed viral sequence reads totaling 30.3 Mb, with an average length of 2,305 kb and a maximum read length of 32,328 bp.

Paired-end Illumina reads and Minion reads were *de novo* assembled using the hybridSPAdes algorithm (Antipov et al., 2016). We obtained two scaffolds representing a total size of 1,595,395 bp with an average read coverage of 121-fold for the first scaffold and 102-fold for the second.

Genome Analysis

Gene predictions were computed using Prodigal (Hyatt et al., 2010). We deleted 49 predicted proteins of fewer than 50 amino acids, or which had abnormal tri-dimensional folding (from 50 to 99 amino acids), as detected by Phyre2 (Kelley et al., 2015). Protein blast was performed against the non-redundant (nr) protein database. Annotation was performed using a combination of Interpro https://www.ebi.ac.uk/interpro/search/sequence-search version 69.0, a CD-search tool online (Marchler-Bauer and Bryant, 2004) and delta-blastp (Boratyn et al., 2012). tRNA prediction was computed online (Lowe and Chan, 2016) with the eukaryotic model parameter.

Comparative Genomics

Orthofinder (v2.4; Emms and Kelly, 2015) was used to infer orthogroups from published genomes of representatives of all established Mimiviridae lineages and estimated high-quality giant virus metagenomes assembled genomes affiliated with the Klosneuvirinae. Further, Interproscan (v5.0; Jones et al., 2014) was employed to assign PFAM-A domains (v33.0; El-Gebali et al., 2019) to giant virus proteins. GenoplotR (Guy et al., 2010) was used to visualize genome synteny of conserved blocks that were inferred with Mauve (v2.4; Darling et al., 2004) in pairwise comparisons between genomes of selected members of the Mimiviridae and Fadolivirus.

Genome Deposition Into Data Repository

The two scaffolds of Fadolivirus IHUMI-VV54 are available on the NCBI website under accession numbers MT418680.1 and MT418681.1.

Phylogenetic Analysis—Species Tree

To infer the phylogenetic position of Fadolivirus in the Mimiviridae, five conserved NCLDV proteins (Yutin et al., 2009) were selected: DNA polymerase elongation subunit family B (NCVOG0038), D5-like helicase-primase (NCVOG0023), packaging ATPase (NCVOG0249), and DNA or RNA helicases of superfamily II (NCVOG0076) and Poxvirus Late Transcription Factor VLTF3-like (NCVOG0262), and identified with hmmsearch (version 3.1b2, hmmer.org). Protein sequences were extracted and aligned using mafft-lins (Katoh and Standley, 2016). Gapped columns in alignments (<10% sequence information) and columns with low information content were removed from the alignment with trimal (v1.4; Capella-Gutiérrez et al., 2009). Protein alignments were then concatenated to a

supermatrix and a species tree of the Mimiviridae was built with IQ-tree (v2.03) with LG + F + R6 (Nguyen et al., 2015). The species tree was then visualized using iTOL v6 online (Letunic and Bork, 2016).

Phylogenetic Analysis - Protein Trees

Phylogenetic analyses for ATPases were conducted as follows. Blastp was used to find 39 close homologous proteins, and amino acid sequences were then aligned using the MUSCLE program (Edgar, 2004). The Mega 6.0 software was used with standard parameters using the maximum likelihood method with 1,000 bootstrap replicates and the Jones–Taylor–Thornton model for amino acid substitution. Phylogenetic trees were then visualized using iTOL v6 online (Letunic and Bork, 2016).

RESULTS

Isolation of a Novel Giant Virus and Host Range Test

Co-culture steps were performed on 64 samples on the amoeba *Vermamoeba vermiformis* strain CDC19 (Boudjemaa et al., 2020). Cell lysis was observed using an inverted optical microscope, with obvious lysis in well 54. Flow cytometry allowed us to detect a single novel population in our gating strategy (Khalil et al., 2016), with a high value of fluorescence compared to the side scatter fluorescence (**Figure 1**). We tried to infect various potential alternate hosts with this virus, specifically *Vermamoeba vermiformis* strain chuv172, *Acanthamoeba castellanii* strain Neff, *Acanthamoeba polyphaga* strain Linc-AP1, *Acanthamoeba royreba* strain Oak ridge, *Dictyostelium discoideum*, the flagellated alga *Tetraselmis* sp., and the protist *Cafeteria roenbergensis*. Fadolivirus was only able to infect *Vermamoeba vermiformis* strain CDC19, suggesting a narrow host range. No cytopathic effects were observed using optical microscopy for any of these potential hosts.

Ultrastructure of Fadolivirus and Replication Cycle

Fadolivirus particles are icosahedral with a diameter of about 300 nm ($n = 21$) on ultrathin sections. Multiple layers of high-density electrons (proteinaceous) were visible comparable to those described for Mimivirus, Yasminevirus and Bodo saltans virus (Deeg et al., 2018; Bajrai et al., 2019). The replication cycle of Fadolivirus showed classical NCLDV stages of infection and replication in *V. vermiformis* and small-sized fibrils were sometimes distinguishable in the periphery of the viral particles, as observed in Yasminevirus. Briefly, the replication cycle begins

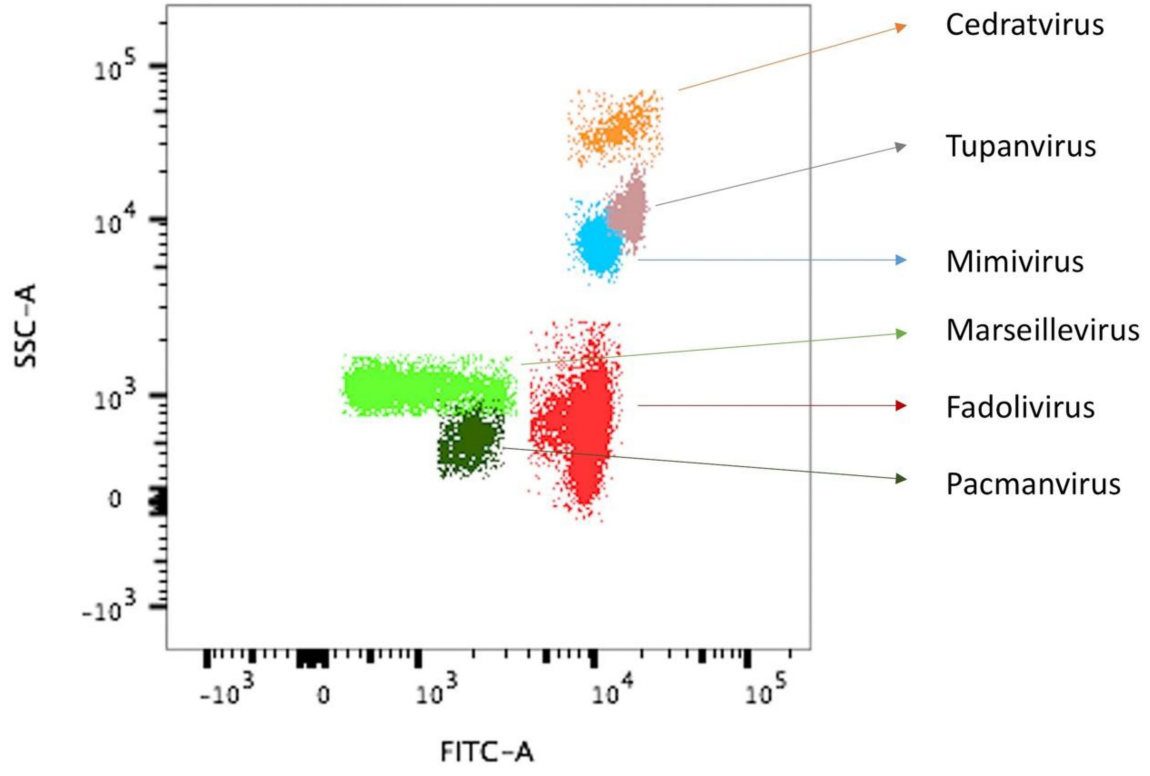


FIGURE 1 | Observation of virus isolate VV54 sample on flow cytometry superimposed with that of other known viral isolates. Flow cytometry dot plot showing different viral profiles, in SSC (side scatter) *versus* FITC (Fluorescein isothiocyanate) SYBR green DNA contents, distinguishable at the levels of capsid shape (optical properties) and DNA components. For the sample VV54, the single viral population (Fadolivirus) is shown in red. For comparison, gates from previous isolates of other viral strains are shown.

with the virus entry by phagocytosis. Next, particles escape the phagosomal process. DNA delivery occurs in the amoeba cytoplasm *via* the apex of the icosahedral virus (**Figures 2A,B**). After cytoplasmic liberation of the virus, an eclipse phase is observed. After 12 h, viral factories are assembled within some amoeba (**Figure 2C**) and 16 h post-infection (**Figure 2D**), nearly all cells present a well-established viral factory. While the Fadolivirus factory resembles that described for Mimivirus (Suzan-Monti et al., 2007), we found a singular element in the filling of the capsid. Indeed, after the synthesis of the icosahedral structure by the factory, the pseudo-ovoidal structure becomes increasingly dense and round inside this neo-formed capsid (**Figures 2F,G**). This step seems to be different compared to observations in Mimivirus. Finally, this gives the particle its mature aspect with a highly dense core. At 20 h post-infection (**Figure 2E**) and 24 h post-infection (**Figure 2H**), the host cells' cytoplasm is fully occupied by newly synthesized virions. Cell burst begins 24 h post-infection and is completed 36 h after the viral infection.

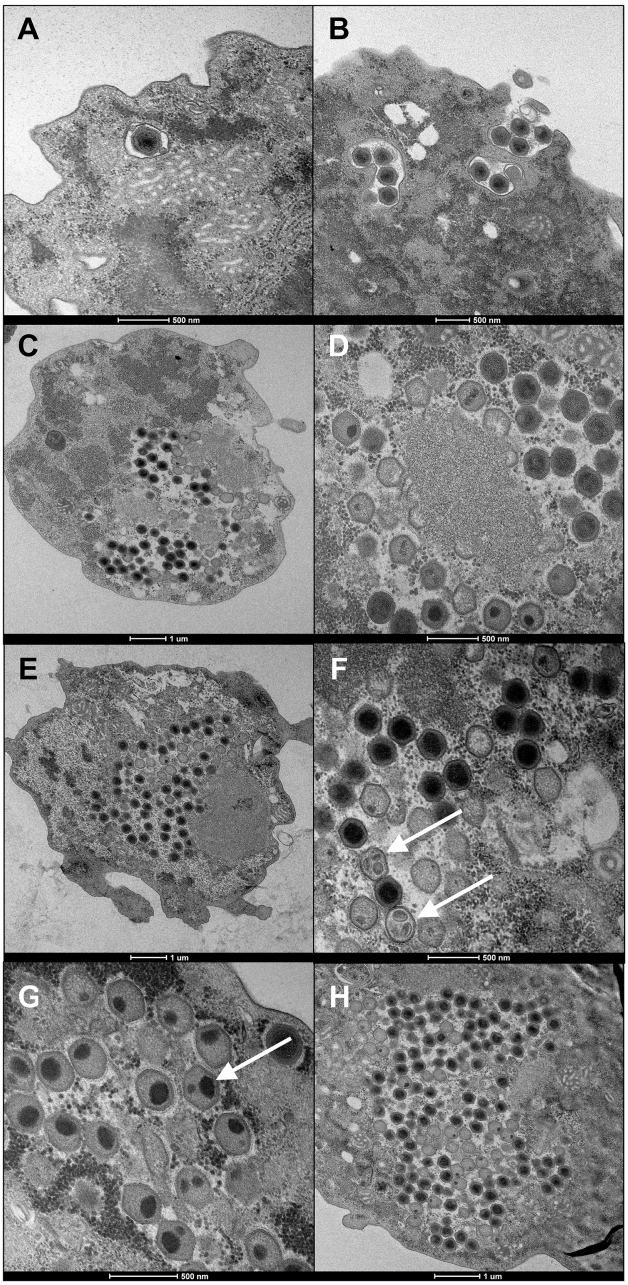


FIGURE 2 | Ultrathin section of the replication cycle of Fadolivirus IHUMI-VV54. Scale bars are indicated under each panel. **(A, B)** were observed during viral entry at 0 h and two hours post-infection. **(C)** Represents a section of *Vermamoeba vermiformis* 12 h post-infection. **(D)** High magnification of a viral factory 16 h post-infection. (E) New virus produced 20 h post-infection. **(F, G)** White arrows indicate icosahedral capsid with ovoid points intra-capsidic localization. **(H)** Intra-cytoplasmic viruses observed 24 h post-infection.

Genomic Features of Fadolivirus

The Fadolivirus genome assembled into two scaffolds with a total length of 1,595,395 bp (with 1,573,504 bp for scaffold 1 and 21,891 bp for scaffold 2; **Table 1**), encoding 1,452 genes of which 1,428 were on the larger scaffold, here referred to as the major scaffold. A megablast of the major scaffold against the nucleotide collection in the NCBI database revealed matches to Indivirus, Klosneuvirus, Catovirus, and Tupanviruses. A whole genome alignment showed that numerous blocks were conserved between Fadolivirus, Klosneuvirus KNV1 and Indivirus ILV1 (**Figure 3**). However, when aligned to the genome of Bodo saltans virus, the proportion of shared sequences decreased notably (data not shown).

A blastp comparison of Fadolivirus proteins to the NCBI nonredundant database (nr) revealed the presence of 256 ORFans without any matches in nr (≈17,6% of all predicted proteins), 552 hits with Klosneuvirus KNV1 (≈38% of all predicted proteins), 223 with Indivirus ILV1 (≈15,4% of all predicted proteins), 24 with Catovirus CTV1, 21 with Hokovirus HKV1 and 24 with Tupanvirus strain deep ocean, 23 with Tupanvirus soda lake, and only 59 with other viruses. We also obtained 111 best hits to eukaryotic and 151 to bacterial and archaeal proteins. Based on the results of this analysis, it becomes evident that the gene content of Fadolivirus is most similar to that of Klosneuvirus. The Fadolivirus genome encodes 66 tRNAs that belong to 19 different types and represent 35 different anticodons (**Table 1; Supplementary Material S1**). Twenty-five were described for the Klosneuvirus KNV1 genome and only one for Indivirus ILV1 (**Table 1**). Similar to what has been described for other Klosneuvirinae (Schulz et al., 2017, 2018), Fadolivirus encoded a comprehensive set of translation system components consisting of various translation initiation factors (eIF4G, F, eIF-5A, B for example), elongation factors (aef-2), and a termination factor (Rho domain).

As previously observed in members of the Klosneuvirinae and in particular in Bodo saltans virus, there were numerous gene duplication events (102 clusters with than two proteins) in the Fadolivirus genome of which only 18 clusters had predicted functions. Thirteen of the duplicated genes were homologous to the capsid protein. More detailed analyses revealed that two ORFs encoded a putative major capsid

TABLE 1 | Morphological and genome characteristics of members of selected Klosneuvirinae and Tupanviruses.

Virus	Fadolivirus IHUMI	Yasminevirus	Klosneuvirus	Catovirus	Hokovirus	Indivirus	Bodo saltans virus	Tupanvirus soda lake	Tupanvirus Deep ocean
Morphological features	Icosahedral capsid	Icosahedral capsid	Unk	Unk	Unk	Unk	Icosahedral capsid	Icosahedral capsid with a tail	Icosahedral capsid with a tail
Genome size (Mbp)	1.59	2.12	1.57	1.53	1.33	0.86	1.38	1.43	1.51
GC content (%)	27.10	40.2	28.6	26.4	21.4	26.6	25.3	29.4	29.1
tRNA	66	70	25	3	0	1	0	67	70
Scaffolds	2	2	16	2	5	18	1	1	1
CDS	1,452	1,541	1,545	1,427	1,022	744	1,207	1,276	1,359
aaRS	23*	20	19	15	3	10	(3*+2)	20	20
CP	11	5	9	Und	Und	Und	4	3	3

*Unk, Unknown; CP, capsid protein; Und, Undetermined. homologous proteins of capsid proteins.*3 in Bodo saltans virus, three aaRS are suggested to represent pseudogenes. *a Glutaminyl-tRNA synthetase is split into two proteins by an HNH endonuclease.*

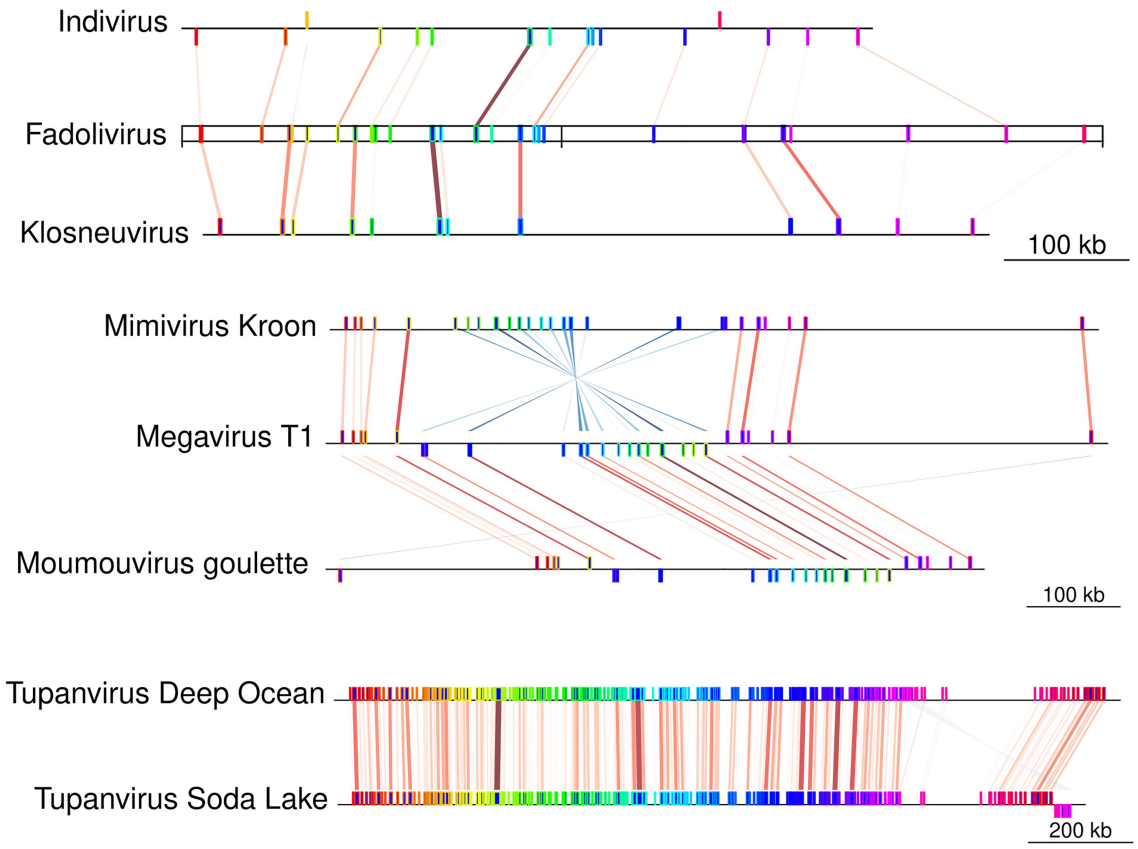


FIGURE 3 | Block synteny comparison between Fadolivirus and other Klosneuvirinae genomes. Conserved genomic regions (Mauve conserved blocks) between Fadolivirus and its closest relatives, Klosneuvirus and Indivirus. Contigs in Indivirus and Klosneuvirus have been reordered based on the Fadolivirus genome. For comparison, conserved regions are also shown between selected members of the proposed subfamily Megamimivirinae, including Tupanviruses.

protein (MCP) split by 35 nucleotides and six others were predicted to represent other MCP (**Supplementary Material S1**). The in-depth *in silico* analysis further led to the exclusion of a putative capsid which was predicted as a probable Zinc finger (NCBI QKF94282.1) domain containing protein. No other *Klosneuvirinae* genome encodes that many (11) capsid proteins. Other conserved NCLDV hallmark genes (Koonin and Yutin, 2010) were also present in multiple copies in the Fadolivirus genome, including the protein predicted to encode the VV-D5 helicase homologues and the A32 packaging ATPase and mRNA capping enzyme with three and four copies, respectively.

In contrast to other members of the *Klosneuvirinae*, Fadolivirus encodes proteins that may play a role in modulation of host cell processes (as for BAG domain-containing protein), vesicular transport as the soluble N-ethylmaleimide-sensitive factor (NSF), pyrimidine metabolism (dihydroorotate dehydrogenase 1A, orotidine 5′-phosphate decarboxylase), transcription and catabolism (putative transcriptional regulator ICP4, regulator of protease activity HflC), in the vitamin K metabolism (vitamin K epoxide reductase and vitamin K dependent gamma-carboxylases (VKGC)), efflux protein (small multidrug resistance protein (multidrug transporter EmrE), in the channel formation (major intrinsic protein, and two predicted proteins in the peptidoglycan synthesis (D-ala-D-ala ligase).

We also retrieved a V21 virophage-like protein (initially described in the Sputnik virophage) that had recently been suggested to play a similar function to transcription factor, as found in other giant virus genomes such as Orpheovirus IHUMI-LCC2 (Andreani et al., 2018). However, no virophage was isolated together with Fadolivirus. Fadolivirus encodes a homolog of Cyclophilin, a chaperone with a role in protein folding (Barik, 2018), yet its role in the Fadolivirus biology remains currently unclear. In the *Acanthamoeba polyphaga Mimivirus*, phosphodiesterase activity was not found (Thai et al., 2008). The Fadolivirus genome also contains three genes annotated as cytochromes (one as 5b, two as P450) and two probable lactonases. Cytochrome has recently been shown to have a broad distribution in DNA viruses and may be involved in cholesterol/sterol synthesis (Lamb et al., 2019).

Fadolivirus Genes Involved in Energy Metabolism

The Fadolivirus genome annotation revealed multiple sequences annotated as putative ATPase subunits and proteins associated with electron transfer with a blue copper enzyme and a putative cytochrome 5b (see previous section). This cytochrome sequence is a probable nitrate reductase (NADH) protein (Chamizo-Ampudia et al., 2017) in the Fadolivirus genome. We identified six proteins associated with proton pumps containing ATPase domains. However, these genes were not co-located in the genome of Fadolivirus. A more detailed phyre2 analysis of these 6 proteins revealed that two of them were indeed predicted as v-type proton ATPase subunit D (QKF94710.1) with 100% confidence and 28% identity with the Yeast V-ATPase state 3 (PDB database 3J9V), and with the other states 1 and 2. This protein could be a vacuolar proton pump hydrolyzing the ATP to ADP (Balakrishna et al., 2015; Zhao et al., 2015). For the protein QKF93735.1, it presents known homologies with vacuolar ATP synthase subunit A by DELTA-blast. This potential subunit A is shared with other giant viruses in Klosneuvirinae, and the predicted subunit D is divergent compared to the eukaryotes (**Figure 4**).

Genomic Diversity Across Klosneuvirinae and Phylogenetic Position of Fadolivirus

In the species tree of the Klosneuvirinae, Fadolivirus branched in a monophyletic clade with Klosneuvirus, Barrevirus and Indivirus (**Figure 5**). This result corresponded well with the complimentary best hit analysis described above. Genomes of members of this clade have a similar GC content of around 30% and genomes of Fadolivirus and Klosneuvirus express a similar genome size with approximately 1.5 Mb, while Indivirus and Barrevirus have much smaller genomes of under 1 Mb. Yasminevirus, another isolated member of the Klosneuvirinae branches divergent from Fadolivirus in the species tree and groups together with Dasosvirus. The third isolate of the Klosneuvirinae, Bodo saltans virus, groups in its own distinct clade. Genomic features and hallmarks that are found in Fadolivirus and Yasminevirus but not in Bodo saltans virus are aaRS. Clusters of orthologous proteins make it possible to observe significant diversity between Mimiviridae (**Figure 6**). There were 12 protein families that are almost exclusively found in the Klosneuvirus/Fadolivirus clade and 4 families that are widespread in the Klosneuvirinae but absent in members of the Klosneuvirus/Fadolivirus clade. Seventy-three orthologs are exclusively shared between Klosneuvirus and Fadolivirus; however, only 20 orthologs are exclusively shared between Fadolivirus, Klosneuvirus and Yasminevirus.

DISCUSSION

Fadolivirus is a new viral isolate with up to 1.59 Mbp, belonging to the sub-family Klosneuvirinae. Its replicative cycle in *V. vermiformis* is comparable to other NCLDVs that have been described, such as Yasminevirus (Bajrai et al., 2019). Fadolivirus shows significant proximity to Klosneuvirus KNV1, virus described by a previous metagenomic study (Schulz et al., 2017), by phylogenetic analysis and by genomic comparison. However, even though these metagenomes assembled viral genomes were of estimated high completeness and low contamination, their hosts and the capacity of each virus to have multiple hosts often remain elusive. Co-occurrence study in Tara Oceans tried to identify probable hosts but their analyses identified families of virus with their probable host in Eukaryotes such as the dominance of Mimiviridae in relation with Alveolata, Opisthokonta, Rhizaria, and Stramenopiles (Meng et al., 2021). Further, based on detection of horizontal gene transfer events in members of the Klosneuvirinae have been suggested to infect Anthoathecata, fungi, arthropods, and various protists including slime molds (Schulz et al., 2020). Indeed, at this time, among amoeba only Tupanviruses have been shown to infect various protists with a similar pattern in *A. castellanii* and in *V. vermiformis* (Abrahão et al., 2018; Silva et al., 2019). Fadolivirus, contrary to earlier predictions, seemed to be rather host specific as it was only able to replicate in *V. vermiformis* (strain CDC19) and not in the *Vermamoeba* strain chuv172. Fadolivirus was also unable to infect any of the other diverse protists that we used in our infection experiments. However, our data are limited to the 8 different protists that were used in the infection experiments. Fadolivirus presents a near complete translation system similar to the ones described in Klosneuvirus (Schulz et al., 2017) and Tupanvirus (Abrahão et al., 2018). Moreover, Fadolivirus has additional copies of tRNA synthetases

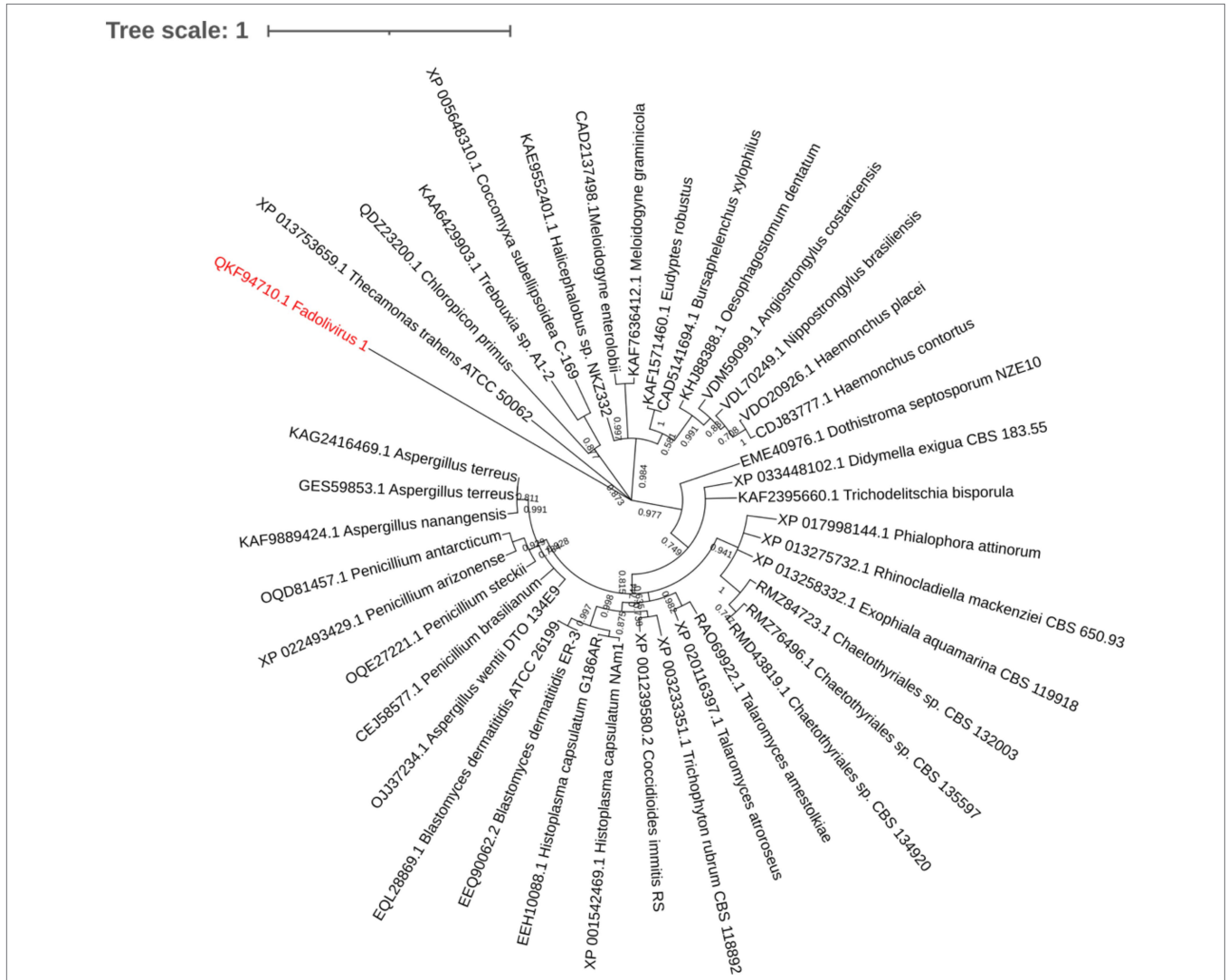


FIGURE 4 | Maximum-likelihood representation built with the Jones–Taylor–Thornton method with 1,000 bootstrap replication based on 286 aligned positions. 39 eukaryote sequences were downloaded as best hit with ATPase from Fadolivirus.

(Asx, Glutaminyl and Ile-lysidine tRNAsynthetases). However, a histidine tRNA synthetase was not found in its genome. It remains unclear whether this absence could lead to a restricted host range that we observed for Fadolivirus. Further analyses will be necessary to provide a more comprehensive overview of potential hosts for this virus. In particular, virus-encoded transcription factors have been proven as being key proteins in members of the *Mimiviridae* during infection of their amoeba host (Bekliz et al., 2018). The presence of translation system genes in Mimiviruses, Tupanviruses, Cafeteria roenbergensis virus, and Klosneuvirinae including Yasminevirus and Bodo saltans virus and combined with the paucity of these genes in the common ancestor of these viruses highlights that such an expanded genetic complement was acquired multiple times and independently in a lineage-specific style of evolution (Schulz et al., 2017; Koonin and Yutin, 2018; Bäckström et al., 2019).

The surprising features of this virus reside in the presence of predicted subunits A and D of ATPase. The first description of vacuolar (H+)-ATPase highlights the close proximity of V and F-type of ATPase D in their catalytic function (Nelson et al., 1995). Recently, during a giant viral infection, it was shown that the shift of the carbonaceous biomass of the host concerned 6–17% for viral production, probably due to a Cedratvirus (Kördel et al., 2021). This observation underlined the fact that viral production requires a greater production of energy by the infected cells. Further, active glycolysis and tricarboxylic acid cycles were identified by metagenomic studies (Moniruzzaman et al., 2020a; Ha et al., 2021) and some proteins were also found in Pandoravirus (Aherfi et al., 2020) and could have a confirmed functional activity such as the isocitrate dehydrogenase protein. Concerning the predicted ATPase subunit A, it is also present in some Klosneuvirinae. We could hypothesize

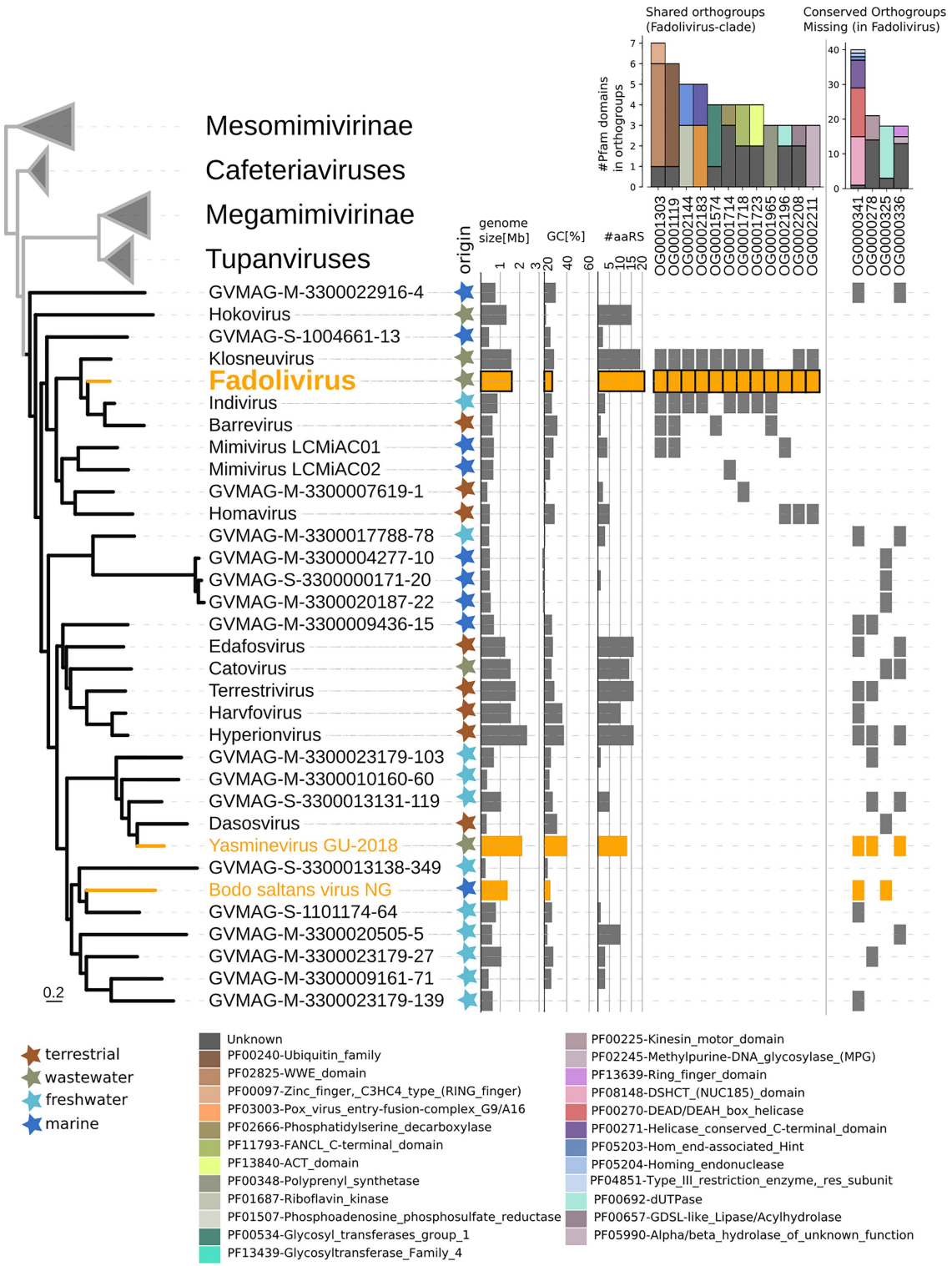


FIGURE 5 | Phylogenetic position of Fadolivirus in the Mimiviridae and genome features. Maximum-likelihood phylogenetic tree (IQ-tree LG + F + R6) of the Mimiviridae inferred from a concatenated protein alignment of five core nucleocytoplasmic virus orthologous genes (NCVOGs; Yutin et al., 2009). Branches in orange represent viral isolates, and branches in black represent previously published metagenome-assembled genomes. The tree is rooted at the proposed viral subfamily Mesomimivirinae. Colored stars next to the tree indicate environmental origin. Bars next to the tree show genome features such as genome size, GC content and total number of proteins that contain aminoacyl-tRNA synthetase (aaRS)-related Pfam domain. Further, the presence/absence matrix next to the tree indicates protein families that

(Continued)

FIGURE 5 | are present in Fadolivirus and its closest relatives but absent in other members of the Mimiviridae; and protein families that are present in more than 30% of viral genomes in the Mimiviridae but not in Fadolivirus and its closest relatives. Stacked bars show the total count and annotation of pfam domains present in each of these protein families.

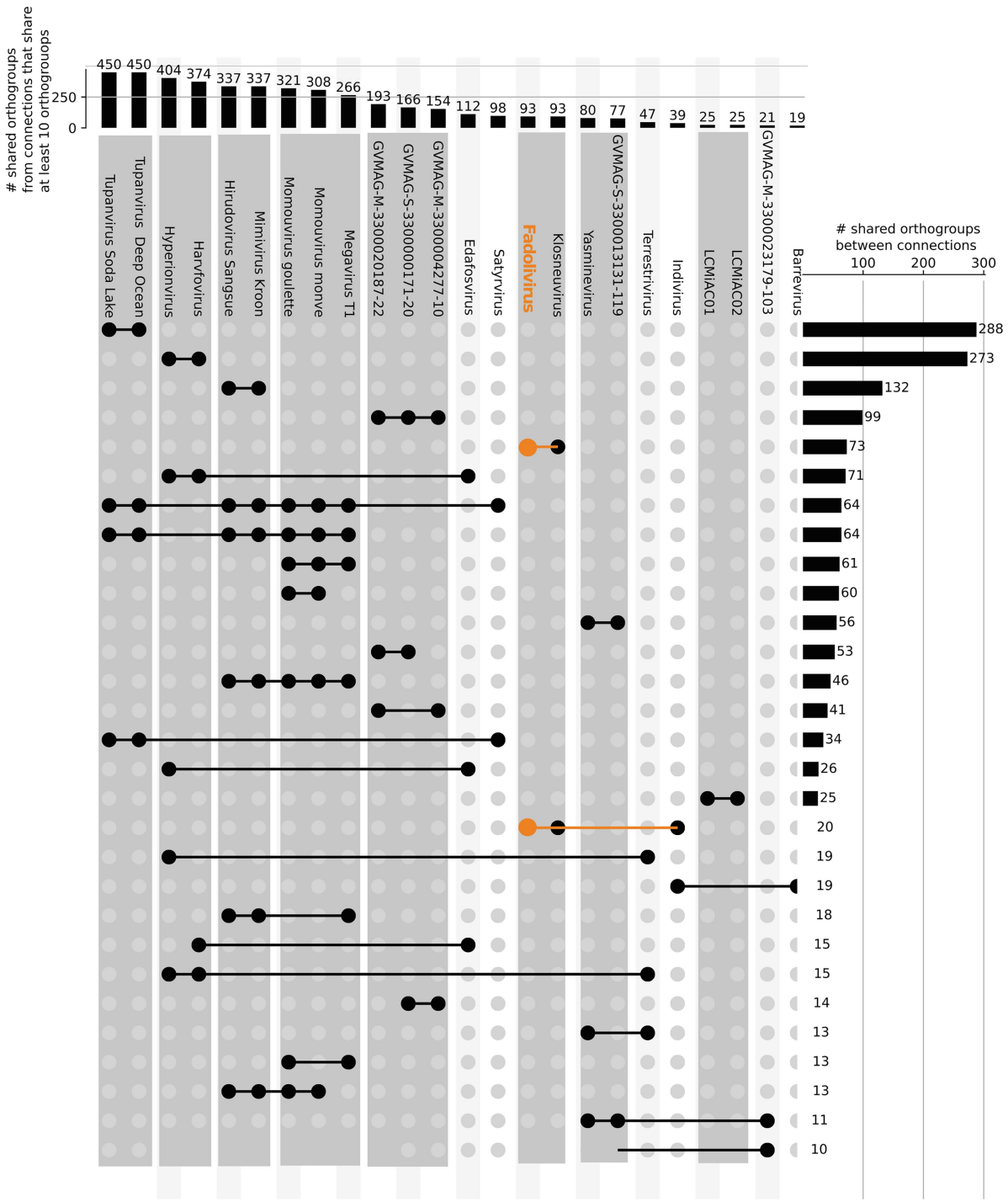


FIGURE 6 | Orthologous genes shared by viruses in the *Klosneuvirinae* and some *Mimiviridae*. Connected lines highlight viruses sharing at least 10 orthologous genes. The bars at the top panel show the total number of shared orthologs between viruses that share at least 10 orthogroups. The bars in the right panel indicate the total number of shared orthogroups between visualized groups of viral genomes (connected by lines in the same row).

that the presence of a more complete system for energy generation system in the ancestor of the *Klosneuvirinae* indicates that the involved genes have been lost in some viral lineages, or those sequences could be highly divergent in ways that are currently not identified by our bioinformatics comparisons. However, for the subunit D this protein is only present for Fadolivirus and could have been acquired from eukaryotes. Nevertheless, in this study we did not perform transcriptomic and proteomic analyses on these putative ATPase subunits; this constitutes a limitation to demonstrate their functional activity.

Meanwhile, novel viruses have been described and appeared as smaller DNA viruses among NCLDV with high speciation to their hosts (Boratto et al., 2020; Subramaniam et al., 2020). Their descriptions bring new opportunities to understanding the diversity and evolution across *Nucleocytoviricota,* and energy systems genes such as the ones observed in Fadolivirus would appear to play a major role during infection by giant viruses.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/**Supplementary Material**.

AUTHOR CONTRIBUTIONS

JA, FS, and BL designed the study and experiments. JA performed cytometry acquisition. JA and FDP performed electron microscopy. JA, FS, and AL performed bioinformatics analyses. JA, FS, TW, and BL wrote the manuscript. All authors approved the final manuscript.

FUNDING

This work was supported by the French Government under the "Investissements d'avenir" (Investments for the Future) programme managed by the Agence Nationale de la Recherche (ANR, fr: National Agency for Research) (reference: Méditerranée Infection 10-IAHU-03). Parts of this study were performed by the US Department of Energy Joint Genome Institute, a DOE Office of Science User Facility and made use of resources of the National Energy Research Scientific Computing Center, both supported by the Office of Science of the US Department of Energy under Contract no. DE-AC02–05CH11231.

ACKNOWLEDGMENTS

The authors would like to thank Claire Andréani for her help with the first English correction.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2021.719703/full#supplementary-material

REFERENCES

Abrahão, J., Silva, L., Silva, L. S., Khalil, J. Y. B., Rodrigues, R., Arantes, T., et al. (2018). Tailed giant Tupanvirus possesses the most complete translational apparatus of the known virosphere. *Nat. Commun.* 9, 1–12. doi: 10.1038/s41467-018-03168-1

Aherfi, S., Andreani, J., Baptiste, E., Oumessoum, A., Dornas, F. P., Andrade, A. C. D. S. P., et al. (2018). A large open Pangenome and a small Core genome for Giant Pandoraviruses. *Front. Microbiol.* 9:1486. doi: 10.3389/fmicb.2018.01486

Aherfi, S., Belhaouari, D. B., Pinault, L., Baudoin, J.-P., Decloquement, P., Abrahão, J., et al. (2020). Tricarboxylic acid cycle and proton gradient in Pandoravirus massiliensis: is it still a virus? *bioRxiv* [Preprint]. doi: 10.1101/2020.09.21.306415

Andreani, J., Aherfi, S., Bou Khalil, J. Y., Di Pinto, F., Bitam, I., Raoult, D., et al. (2016). Cedratvirus, a double-Cork structured Giant virus, is a distant relative of Pithoviruses. *Viruses* 8:300. doi: 10.3390/v8110300

Andreani, J., Bou Khalil, J. Y., Sevvana, M., Benamar, S., Di Pinto, F., Bitam, I., et al. (2017). Pacmanvirus, a new giant icosahedral virus at the crossroads between Asfarviridae and Faustoviruses. *J. Virol.* 91, e00212–e00217. doi: 10.1128/JVI.00212-17

Andreani, J., Khalil, J. Y. B., Baptiste, E., Hasni, I., Michelle, C., Raoult, D., et al. (2018). Orpheovirus IHUMI-LCC2: a new virus among the Giant viruses. *Front. Microbiol.* 8:2643. doi: 10.3389/fmicb.2017.02643

Antipov, D., Korobeynikov, A., McLean, J. S., and Pevzner, P. A. (2016). hybridSPAdes: an algorithm for hybrid assembly of short and long reads. *Bioinformatics* 32, 1009–1015. doi: 10.1093/bioinformatics/btv688

Antwerpen, M. H., Georgi, E., Zoeller, L., Woelfel, R., Stoecker, K., and Scheid, P. (2015). Whole-genome sequencing of a pandoravirus isolated from keratitis-inducing acanthamoeba. *Genome Announc.* 3, e00136–e00115. doi: 10.1128/genomeA.00136-15

Bäckström, D., Yutin, N., Jørgensen, S. L., Dharamshi, J., Homa, F., Zaremba-Niedwiedzka, K., et al. (2019). Virus genomes from Deep Sea sediments expand the ocean Megavirome and support independent origins of viral gigantism. *MBio* 10, e02497–e02418. doi: 10.1128/mBio.02497-18

Bajrai, L. H., Mougari, S., Andreani, J., Baptiste, E., Delerce, J., Raoult, D., et al. (2019). Isolation of Yasminevirus, the first member of Klosneuvirinae isolated in coculture with Vermamoeba vermiformis, demonstrates an extended arsenal of translational apparatus components. *J. Virol.* 94, e01534–e01519. doi: 10.1128/JVI.01534-19

Balakrishna, A. M., Basak, S., Manimekalai, M. S. S., and Grüber, G. (2015). Crystal structure of subunits D and F in complex gives insight into energy transmission of the eukaryotic V-ATPase from Saccharomyces cerevisiae. *J. Biol. Chem.* 290, 3183–3196. doi: 10.1074/jbc.M114.622688

Barik, S. (2018). A family of novel Cyclophilins, conserved in the Mimivirus genus of the Giant DNA viruses. *Comput. Struct. Biotechnol. J.* 16, 231–236. doi: 10.1016/j.csbj.2018.07.001

Bekliz, M., Azza, S., Seligmann, H., Decloquement, P., Raoult, D., and La Scola, B. (2018). Experimental analysis of Mimivirus translation initiation factor 4a reveals its importance in viral protein translation during infection of Acanthamoeba polyphaga. *J. Virol.* 92, e00337–e00318. doi: 10.1128/JVI.00337-18

Bertelli, C., Mueller, L., Thomas, V., Pillonel, T., Jacquier, N., and Greub, G. (2017). Cedratvirus lausannensis-digging into Pithoviridaediversity. *Environ. Microbiol.* 19, 4022–4034. doi: 10.1111/1462-2920.13813

Boratto, P. V. M., Oliveira, G. P., Machado, T. B., Andrade, A. C. S. P., Baudoin, J.-P., Klose, T., et al. (2020). Yaravirus: a novel 80-nm virus infecting Acanthamoeba castellanii. *Proc. Natl. Acad. Sci.* 7:202001637. doi: 10.1073/pnas.2001637117

Boratyn, G. M., Schäffer, A. A., Agarwala, R., Altschul, S. F., Lipman, D. J., and Madden, T. L. (2012). Domain enhanced lookup time accelerated BLAST. *Biol. Direct* 7, 12–14. doi: 10.1186/1745-6150-7-12

Boudjemaa, H., Andreani, J., Bitam, I., and La Scola, B. (2020). Diversity of amoeba-associated Giant viruses isolated in Algeria. *Diversity* 12:215. doi: 10.3390/d12060215

Capella-Gutiérrez, S., Silla-Martínez, J. M., and Gabaldón, T. (2009). trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* 25, 1972–1973. doi: 10.1093/bioinformatics/btp348

Chamizo-Ampudia, A., Sanz-Luque, E., Llamas, A., Galvan, A., and Fernandez, E. (2017). Nitrate reductase regulates plant nitric oxide homeostasis. *Trends Plant Sci.* 22, 163–174. doi: 10.1016/j.tplants.2016.12.001

Colson, P., La Scola, B., Levasseur, A., Caetano-Anollés, G., and Raoult, D. (2017). Mimivirus: leading the way in the discovery of giant viruses of amoebae. *Nat. Rev. Microbiol.* 15, 243–254. doi: 10.1038/nrmicro.2016.197

Darling, A. C. E., Mau, B., Blattner, F. R., and Perna, N. T. (2004). Mauve: multiple alignment of conserved genomic sequence with rearrangements. *Genome Res.* 14, 1394–1403. doi: 10.1101/gr.2289704

Deeg, C. M., Chow, C.-E. T., and Suttle, C. A. (2018). The kinetoplastid-infecting Bodo saltans virus (BsV), a window into the most abundant giant viruses in the sea. *elife* 7:e33014. doi: 10.7554/eLife.33014

Edgar, R. C. (2004). MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinfo.* 5:113. doi: 10.1186/1471-2105-5-113

El-Gebali, S., Mistry, J., Bateman, A., Eddy, S. R., Luciani, A., Potter, S. C., et al. (2019). The Pfam protein families database in 2019. *Nucl. Acids Res.* 47, D427–D432. doi: 10.1093/nar/gky995

Emms, D. M., and Kelly, S. (2015). OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome Biol.* 16, 1–14. doi: 10.1186/s13059-015-0721-2

Forterre, P. (2016). To be or not to be alive: how recent discoveries challenge the traditional definitions of viruses and life. *Stud. Hist. Philos. Sci. Part C: Stud. Hist. Philos. Biol. Biomed. Sci.* 59, 100–108. doi: 10.1016/j.shpsc.2016.02.013

Gaia, M., Benamar, S., Boughalmi, M., Pagnier, I., Croce, O., Colson, P., et al. (2014). Zamilon, a novel Virophage with Mimiviridae host specificity. *PLoS One* 9:e94923. doi: 10.1371/journal.pone.0094923

Geballa-Koukoulas, K., Andreani, J., La Scola, B., and Blanc, G. (2021). The Kaumoebavirus LCC10 genome reveals a unique Gene Strand bias among "extended Asfarviridae". *Viruses* 13:148. doi: 10.3390/v13020148

Geballa-Koukoulas, K., Boudjemaa, H., Andreani, J., La Scola, B., and Blanc, G. (2020). Comparative genomics unveils regionalized evolution of the Faustovirus genomes. *Viruses* 12:577. doi: 10.3390/v12050577

Guy, L., Kultima, J. R., and Andersson, S. G. E. (2010). genoPlotR: comparative gene and genome visualization in R. *Bioinformatics* 26, 2334–2335. doi: 10.1093/bioinformatics/btq413

Ha, A. D., Moniruzzaman, M., and Aylward, F. O. (2021). High transcriptional activity and diverse functional repertoires of hundreds of Giant viruses in a coastal marine system. *bioRxiv* 2021:e0029321. doi: 10.1101/2021.03.08.434518

Hyatt, D., Chen, G.-L., LoCascio, P. F., Land, M. L., Larimer, F. W., and Hauser, L. J. (2010). Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* 11:119. doi: 10.1186/1471-2105-11-119

Jones, P., Binns, D., Chang, H.-Y., Fraser, M., Li, W., McAnulla, C., et al. (2014). InterProScan 5: genome-scale protein function classification. *Bioinformatics* 30, 1236–1240. doi: 10.1093/bioinformatics/btu031

Katoh, K., and Standley, D. M. (2016). A simple method to control over-alignment in the MAFFT multiple sequence alignment program. *Bioinformatics* 32, 1933–1942. doi: 10.1093/bioinformatics/btw108

Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N., and Sternberg, M. J. E. (2015). The Phyre2 web portal for protein modeling, prediction and analysis. *Nat. Protoc.* 10, 845–858. doi: 10.1038/nprot.2015.053

Khalil, J. Y. B., Robert, S., Reteno, D. G., Andreani, J., Raoult, D., and La Scola, B. (2016). High-throughput isolation of Giant viruses in liquid medium using automated flow cytometry and fluorescence staining. *Front. Microbiol.* 7:26. doi: 10.3389/fmicb.2016.00026

Koonin, E. V., Dolja, V. V., Krupovic, M., Varsani, A., Wolf, Y. I., Yutin, N., et al. (2020). Global organization and proposed Megataxonomy of the virus world. *Microbiol. Mol. Biol. Rev.* 84, e00061–e00019. doi: 10.1128/MMBR.00061-19

Koonin, E. V., and Yutin, N. (2010). Origin and evolution of eukaryotic large nucleo-cytoplasmic DNA viruses. *Intervirology* 53, 284–292. doi: 10.1159/000312913

Koonin, E. V., and Yutin, N. (2018). Multiple evolutionary origins of giant viruses. *F1000Research* 7:1840. doi: 10.12688/f1000research.16248.1

Kördel, M., Svenda, M., Reddy, H. K. N., Fogelqvist, E., Arsana, K. G. Y., Hamawandi, B., et al. (2021). Quantitative conversion of biomass in giant DNA virus infection. *Sci. Rep.* 11, 1–12. doi: 10.1038/s41598-021-83547-9

La Scola, B. (2003). A Giant Virus in Amoebae. *Science* 299:2033. doi: 10.1126/science.1081867

La Scola, B., Desnues, C., Pagnier, I., Robert, C., Barrassi, L., Fournous, G., et al. (2008). The virophage as a unique parasite of the giant mimivirus. *Nature* 455, 100–104. doi: 10.1038/nature07218

Lamb, D. C., Follmer, A. H., Goldstone, J. V., Nelson, D. R., Warrilow, A. G., Price, C. L., et al. (2019). On the occurrence of cytochrome P450 in viruses. *Proc. Natl. Acad. Sci.* 116, 12343–12352. doi: 10.1073/pnas.1901080116

Legendre, M., Bartoli, J., Shmakova, L., Jeudy, S., Labadie, K., Adrait, A., et al. (2014). Thirty-thousand-year-old distant relative of giant icosahedral DNA viruses with a pandoravirus morphology. *Proc. Natl. Acad. Sci.* 111, 4274–4279. doi: 10.1073/pnas.1320670111

Legendre, M., Fabre, E., Poirot, O., Jeudy, S., Lartigue, A., Alempic, J.-M., et al. (2018). Diversity and evolution of the emerging Pandoraviridae family. *Nat. Commun.* 9, 1–12. doi: 10.1038/s41467-018-04698-4

Letunic, I., and Bork, P. (2016). Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. *Nucl. Acids Res.* 44, W242–W245. doi: 10.1093/nar/gkw290

Levasseur, A., Andreani, J., Delerce, J., Bou Khalil, J., Robert, C., La Scola, B., et al. (2016). Comparison of a modern and fossil Pithovirus reveals its genetic conservation and evolution. *Genome Biol. Evol.* 8, 2333–2339. doi: 10.1093/gbe/evw153

Lowe, T. M., and Chan, P. P. (2016). tRNAscan-SE on-line: integrating search and context for analysis of transfer RNA genes. *Nucl. Acids Res.* 44, W54–W57. doi: 10.1093/nar/gkw413

Marchler-Bauer, A., and Bryant, S. H. (2004). CD-search: protein domain annotations on the fly. *Nucleic Acids Res.* 32, W327–W331. doi: 10.1093/nar/gkh454

Meng, L., Endo, H., Blanc-Mathieu, R., Chaffron, S., Hernández-Velázquez, R., Kaneko, H., et al. (2021). Quantitative assessment of nucleocytoplasmic large DNA virus and host interactions predicted by co-occurrence analyses. *mSphere* 6, e01298–e01220. doi: 10.1128/mSphere.01298-20

Moniruzzaman, M., Martinez-Gutierrez, C. A., Weinheimer, A. R., and Aylward, F. O. (2020a). Dynamic genome evolution and complex virocell metabolism of globally-distributed giant viruses. *Nat. Commun.* 11, 1710–1711. doi: 10.1038/s41467-020-15507-2

Moniruzzaman, M., Weinheimer, A. R., Martinez-Gutierrez, C. A., and Aylward, F. O. (2020b). Widespread endogenization of giant viruses shapes genomes of green algae. *Nature* 588, 141–145. doi: 10.1038/s41586-020-2924-2

Nelson, H., Mandiyan, S., and Nelson, N. (1995). A bovine cDNA and a yeast gene (VMA8) encoding the subunit D of the vacuolar H(+)-ATPase. *Proc. Natl. Acad. Sci.* 92, 497–501. doi: 10.1073/pnas.92.2.497

Nguyen, L.-T., Schmidt, H. A., and Haeseler, von, A., and Minh, B. Q., (2015). IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* 32, 268–274. doi: 10.1093/molbev/msu300

Philippe, N., Legendre, M., Doutre, G., Couté, Y., Poirot, O., Lescot, M., et al. (2013). Pandoraviruses: amoeba viruses with genomes up to 2.5 Mb reaching that of parasitic eukaryotes. *Science* 341, 281–286. doi: 10.1126/science.1239181

Reteno, D. G., Benamar, S., Khalil, J. B., Andreani, J., Armstrong, N., Klose, T., et al. (2015). Faustovirus, an asfarvirus-related new lineage of giant viruses infecting amoebae. *J. Virol.* 89, 6585–6594. doi: 10.1128/JVI.00115-15

Rodrigues, R. A. L., Andreani, J., Andrade, A. C. D. S. P., Machado, T. B., Abdi, S., Levasseur, A., et al. (2018a). Morphologic and genomic analyses of new isolates reveal a second lineage of cedratviruses. *J. Virol.* 92, 00372–00318. doi: 10.1128/JVI.00372-18

Rodrigues, R. A. L., Mougari, S., Colson, P., La Scola, B., and Abrahão, J. S. (2018b). "Tupanvirus," a new genus in the family Mimiviridae. *Arch. Virol.* 15, 243–247. doi: 10.1007/s00705-018-4067-4

Schulz, F., Alteio, L., Goudeau, D., Ryan, E. M., Yu, F. B., Malmstrom, R. R., et al. (2018). Hidden diversity of soil giant viruses. *Nat. Commun.* 9, 1–9. doi: 10.1038/s41467-018-07335-2

Schulz, F., Roux, S., Paez-Espino, D., Jungbluth, S., Walsh, D. A., Denef, V. J., et al. (2020). Giant virus diversity and host interactions through global metagenomics. *Nature* 578, 432–436. doi: 10.1038/s41586-020-1957-x

Schulz, F., Yutin, N., Ivanova, N. N., Ortega, D. R., Lee, T. K., Vierheilig, J., et al. (2017). Giant viruses with an expanded complement of translation system components. *Science* 356, 82–85. doi: 10.1126/science.aal4657

Shinn, G. L., and Bullard, B. L. (2018). Ultrastructure of Meelsvirus: a nuclear virus of arrow worms (phylum Chaetognatha) producing giant "tailed" virions. *PLoS One* 13:e0203282. doi: 10.1371/journal.pone.0203282

Silva, L. K. D. S., Andrade, A. C. D. S. P., Dornas, F. P., Rodrigues, R. A. L., Arantes, T., Kroon, E. G., et al. (2018). Cedratvirus getuliensis replication cycle: an in-depth morphological analysis. *Sci. Rep.* 8, 1–11. doi: 10.1038/s41598-018-22398-3

Silva, L. C. F., Rodrigues, R. A. L., Oliveira, G. P., Dornas, F. P., La Scola, B., Kroon, E. G., et al. (2019). Microscopic analysis of the Tupanvirus cycle in Vermamoeba vermiformis. *Front. Microbiol.* 10:671. doi: 10.3389/fmicb.2019.00671

Subramaniam, K., Behringer, D. C., Bojko, J., Yutin, N., Clark, A. S., Bateman, K. S., et al. (2020). A new family of DNA viruses causing disease in crustaceans from diverse aquatic biomes. *MBio* 11, e02938–e02919. doi: 10.1128/mBio.02938-19

Suzan-Monti, M., La Scola, B., Barrassi, L., Espinosa, L., and Raoult, D. (2007). Ultrastructural characterization of the giant volcano-like virus factory of Acanthamoeba polyphaga Mimivirus. *PLoS One* 2:e328. doi: 10.1371/journal.pone.0000328

Thai, V., Renesto, P., Fowler, C. A., Brown, D. J., Davis, T., Gu, W., et al. (2008). Structural, biochemical, and in vivo characterization of the first virally encoded cyclophilin from the Mimivirus. *J. Mol. Biol.* 378, 71–86. doi: 10.1016/j.jmb.2007.08.051

Verneau, J., Levasseur, A., Raoult, D., La Scola, B., and Colson, P. (2016). MG-digger: an automated pipeline to search for Giant virus-related sequences in metagenomes. *Front. Microbiol.* 7:428. doi: 10.3389/fmicb.2016.00428

Wilson, W. H., Gilg, I. C., Moniruzzaman, M., Field, E. K., Koren, S., LeCleir, G. R., et al. (2017). Genomic exploration of individual giant ocean viruses. *ISME J.* 15, 1736–1745. doi: 10.1038/ismej.2017.61

Yutin, N., Wolf, Y. I., Raoult, D., and Koonin, E. V. (2009). Eukaryotic large nucleo-cytoplasmic DNA viruses: clusters of orthologous genes and reconstruction of viral genome evolution. *Virol. J.* 6:223. doi: 10.1186/1743-422X-6-223

Zhao, J., Benlekbir, S., and Rubinstein, J. L. (2015). Electron cryomicroscopy observation of rotational states in a eukaryotic V-ATPase. *Nature* 521, 241–245. doi: 10.1038/nature14365

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Advantages of publishing in Frontiers

OPEN ACCESS
Articles are free to read for greatest visibility and readership

FAST PUBLICATION
Around 90 days from submission to decision

HIGH QUALITY PEER-REVIEW
Rigorous, collaborative, and constructive peer-review

TRANSPARENT PEER-REVIEW
Editors and reviewers acknowledged by name on published articles

Frontiers
Avenue du Tribunal-Fédéral 34
1005 Lausanne | Switzerland

Visit us: www.frontiersin.org
Contact us: frontiersin.org/about/contact

REPRODUCIBILITY OF RESEARCH
Support open data and methods to enhance research reproducibility

DIGITAL PUBLISHING
Articles designed for optimal readership across devices

FOLLOW US
@frontiersin

IMPACT METRICS
Advanced article metrics track visibility across digital media

EXTENSIVE PROMOTION
Marketing and promotion of impactful research

LOOP RESEARCH NETWORK
Our network increases your article's readership